UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for
New Economic Thinking

CAMBRIDGE WORKING PAPERS IN ECONOMICS
CAMBRIDGE-INET WORKING PAPERS

Connectors and Influencers

Syngjoo Choi	Sanjeev Goyal	Frédéric Moisan
Seoul National University	University of Cambridge	EM Lyon Business School

Abstract

Classical models of network formation bring out the salience of a `law of the few' property, that is manifest in hub-spoke/star architectures. Existing experiments on network formation find that subjects do not create such networks.

Our paper conducts a network formation experiment on the model of Galeotti and Goyal (2010). The theory predicts that every equilibrium of this game is a `star' network in which the spokes pay for links with a single hub. There are two equilibrium effort configurations: the center makes all the effort (the pure influencer outcome) and the hub makes zero effort (the pure connector outcome). This paper tests these predictions with the help of a new experimental platform with asynchronous activity in continuous time. We vary group size and provision of information of others' payoffs.

Subjects always create networks with specialization in linking. This is consistent with equilibrium prediction. Our second result concerns the interaction of group size and information provision. In a baseline information treatment where subjects only see their own payoffs, they select the pure influencer outcome. By contrast, when we provide information on everyone's payoffs, in large groups, subjects select a pure connector outcome. These behavioural patterns can be accounted for by a decision rule on activity level that combines myopic best response and competition for hub status.

Reference Details

1935	Cambridge Working Papers in Economics
2019/09	Cambridge-INET Working Paper Series
Published	28 March 2019
Revised	29 March 2021
JEL Codes	C92, D83, D85, Z13
Websites	www.econ.cam.ac.uk/cwpe www.inet.econ.cam.ac.uk/working-papers

Connectors and Influencers

Syngjoo Choi* Sanjeev Goyal† Frédéric Moisan‡

March 29, 2021

Abstract

Classical models of network formation bring out the salience of a 'law of the few' property, that is manifest in hub-spoke/star architectures. Existing experiments on network formation find that subjects do not create such networks.

Our paper conducts a network formation experiment on the model of Galeotti and Goyal (2010). The theory predicts that every equilibrium of this game is a 'star' network in which the spokes pay for links with a single hub. There are two equilibrium effort configurations: the center makes all the effort (the pure influencer outcome) and the hub makes zero effort (the pure connector outcome). This paper tests these predictions with the help of a new experimental platform with asynchronous activity in continuous time. We vary group size and provision of information of others' payoffs.

Subjects always create networks with specialization in linking. This is consistent with equilibrium prediction. Our second result concerns the interaction of group size and information provision. In a baseline information treatment where subjects only see their own payoffs, they select the pure influencer outcome. By contrast, when we provide information on everyone's payoffs, in large groups, subjects select a pure connector outcome. These behavioural patterns can be accounted for by a decision rule on activity level that combines myopic best response and competition for hub status.

JEL: C92, D83, D85, Z13.

*Department of Economics, Seoul National University, Seoul 08826, Republic of Korea. Email: syngjooc@snu.ac.kr

†Faculty of Economics and Christ's College, University of Cambridge, Cambridge CB3 9DD, United Kingdom. Email: sg472@cam.ac.uk

‡EM Lyon Business School, GATE UMR 5824, 69130 Écully, France. Email: moisan@em-lyon.com

The paper was supported by the FET-Open Project IBSEN (contract no. 662725), the Keynes Fund and by Creative-Pioneering Researchers Program (Seoul National University), and was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon. We thank participants at a number of seminars and conferences for very helpful comments.

1 Introduction

Large scale social networks are a defining feature of contemporary economy and society. Empirical research suggests that such networks exhibit a *law of the few*: the distribution of links is very unequal.[1] Given the social and economic implications of this inequality, it is important to understand the principles underlying the formation of these networks.

The economic approach to network formation takes the view that networks are created by individuals who compare the costs and benefits of linking. Beginning with the early work of Bala and Goyal (2000) and Jackson and Wolinsky (1996), this idea has been explored in a number of papers on network formation. A high level take away from this literature is that linking activity leads to the 'law of the few'.[2] This result has been the subject of extended experimental investigation: in a test of the Bala and Goyal (2000) model, Falk and Kosfeld (2012) and Goeree, Riedl, and Ule (2009) show that experimental subjects do not create such networks; in a recent paper that tests a model of linking and efforts taken from Galeotti and Goyal (2010), van Leeuwen, Offerman, and Schram (2019) report that the specialization in linking and efforts predicted by the model is not observed in the laboratory. These experimental findings raise a question mark about the validity of these models and more generally about an economic approach to understanding networks.

A common feature of existing experiments is that the number of subjects is small (typically ranging between 4 and 8). Moreover, practically all the experiments require subjects to make simultaneous choices in discrete time. In a real world setting, groups are very large and individuals typically choose effort and linking at different points in time. The individual decision problem is complicated because the attractiveness of links depends on the efforts of individuals *and* also on the efforts by the neighbours of these individuals. As group size grows, these informational requirements become more demanding. So it is unclear if we can extend the findings from the small group experiments to more realistic settings. The work of Berninghaus, Ehrhart, and Ott (2006), Friedman and Oprea (2012) and Goyal, Rosenkranz, Weitzel, and Buskens (2017) suggests that continuous time experiments offer subjects more opportunities for choice and for learning and that they may offer better prospects for convergence to equilibrium than discrete time experiments. Our paper builds on this insight. We develop a new platform for network experiments in which

[1] See Barabsi and Albert (1999), Goyal, Moraga, and van der Leij (2006), and Jackson and Rogers (2007).

[2] See e.g., Hojman and Szeidl (2008); Bramoulle, Galeotti, and Rogers (2016) provide a survey of the large networks literature.

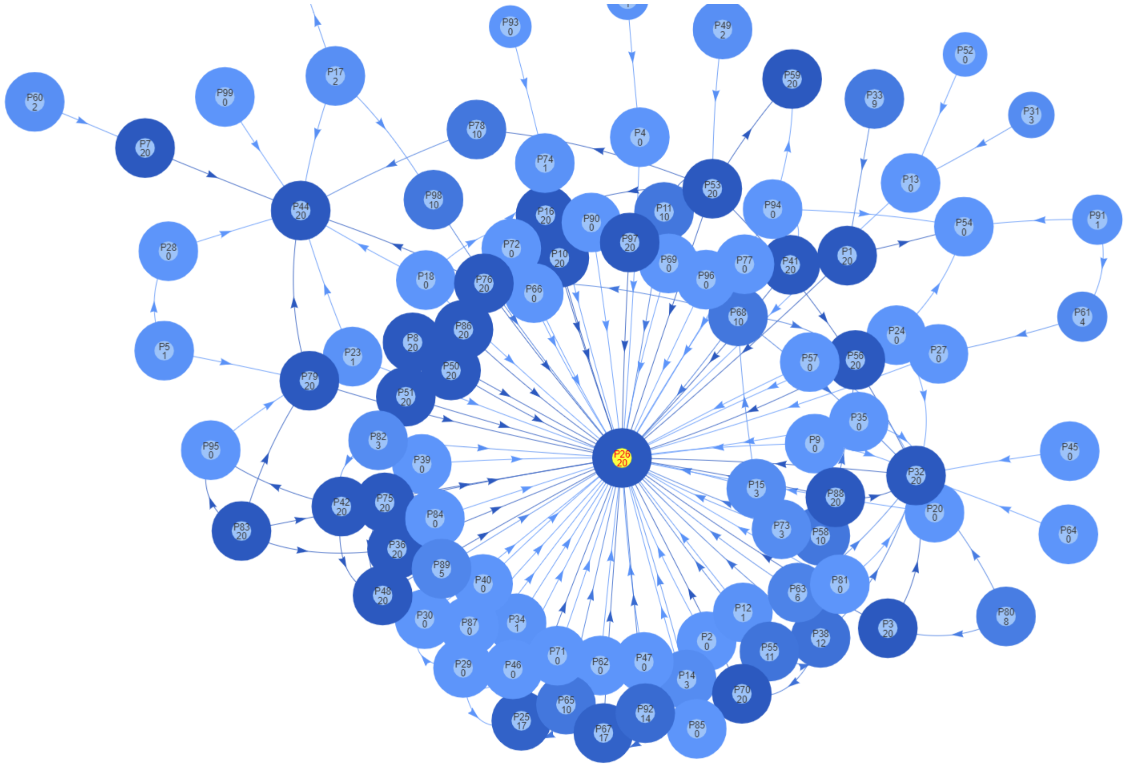
individual choice is asynchronous and takes place in continuous time and also allows for large groups of up to 100 subjects.

The paper reports on an experiment on the model of linking and efforts taken from Galeotti and Goyal (2010).[3] The theory predicts that every (Nash) equilibrium of this game is a 'star' network in which the spokes pay for links with a single hub. There are two equilibrium effort configurations: the center makes all the effort (the pure influencer outcome) and the hub makes zero effort (the pure connector outcome). The goal of the paper is to test the relevance of these predictions in the static environment to outcomes in the environment with realistic opportunities for linking and activity. There are four group sizes $4, 8, 50, 100$ and each of these groups plays the linking and effort game over 6 minutes. There are two information treatments: in the baseline treatment, subjects observe only their own payoffs, while in the payoff information treatment a subject observes the payoffs of everyone. Taken together, we therefore have a 4×2 design. This design enables us to vary the strategic uncertainty and informational load facing subjects and therefore offers a general environment to test the theory.

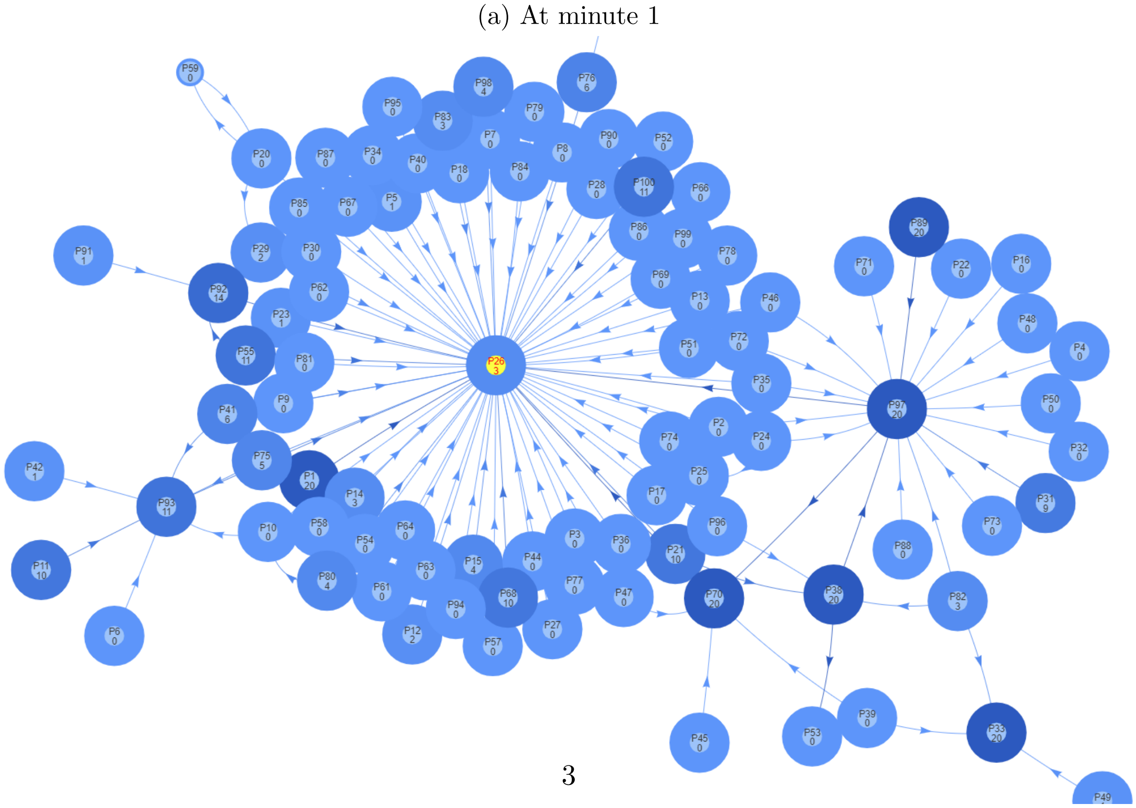
We start with the baseline information treatment. Figures 1 and 2 present snapshots taken from the experiment with a hundred subjects. Initially, at minute 1, subject P26 emerges as a hub with the maximum effort 20. There are other subjects who make maximal effort (such as P97). At minute 3, P26 continues to be a hub but has substantially lowered her effort. Due to this shading of effort, she starts to lose some of her links to subject P97, who has kept her effort at 20. The transition becomes clearer in Figure 2a at the 5 minute mark, when the initial hub subject P26 has lost most of her links to the emerging hub P97. Figure 2b confirms that this transition is stable until the end of the game.

Our first finding is that, in all four group sizes, there is specialization in linking and efforts. This manifests itself in the clearest form in the large groups (as in the 100 subject experiment reported above). Our second finding concerns individual behavior. In all four group sizes, highly connected individuals exert large efforts. In particular, in small groups the efforts of the most connected individual are close to the equilibrium prediction of the static model. By contrast, in the large groups of fifty and hundred subjects the most connected subject chose efforts that are far higher than the equilibrium prediction. As a consequence, in the large groups, the hubs earn less than the peripheral nodes.

[3]A number of paper have explored this framework, see e.g., Baetz (2015), Perego and Yuksel (2016) and Herskovic and Ramos (2020). These models may be seen as combining the two-way linking model of Bala and Goyal (2000) with the public goods model in networks model of Bramoulle and Kranton (2007).

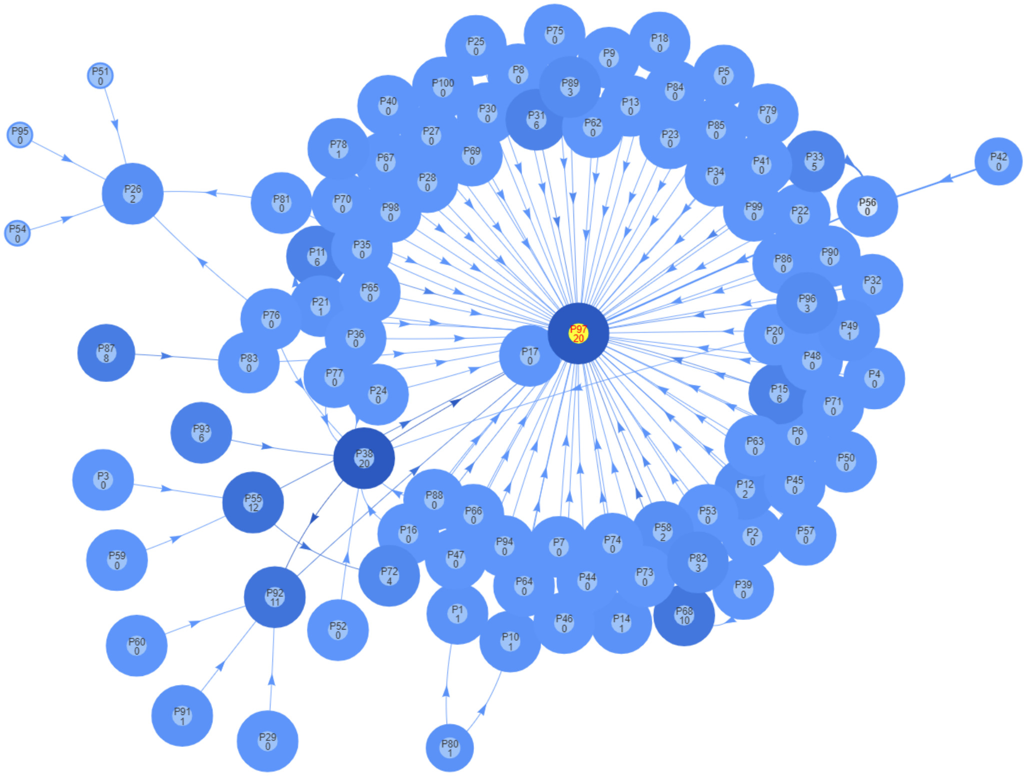


(a) At minute 1

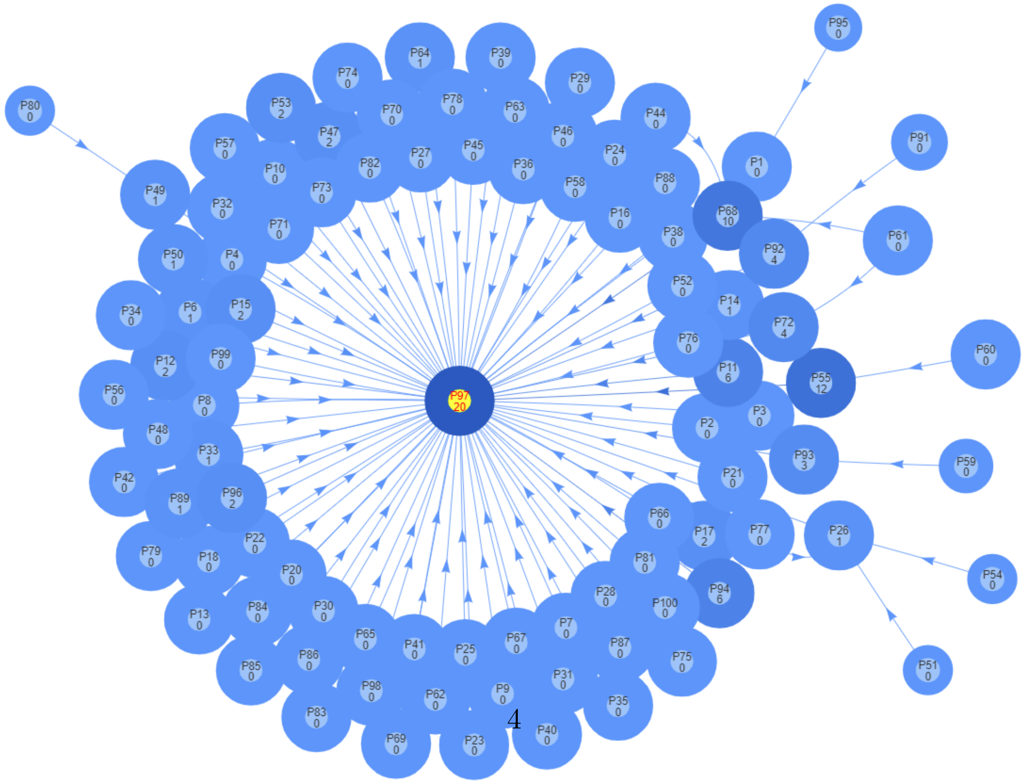


(b) at minute 3

Figure 1: Snap shots of a dynamic game



(a) At minute 5



(b) At minute 6

Figure 2: Snap shots of a dynamic game (cont.)

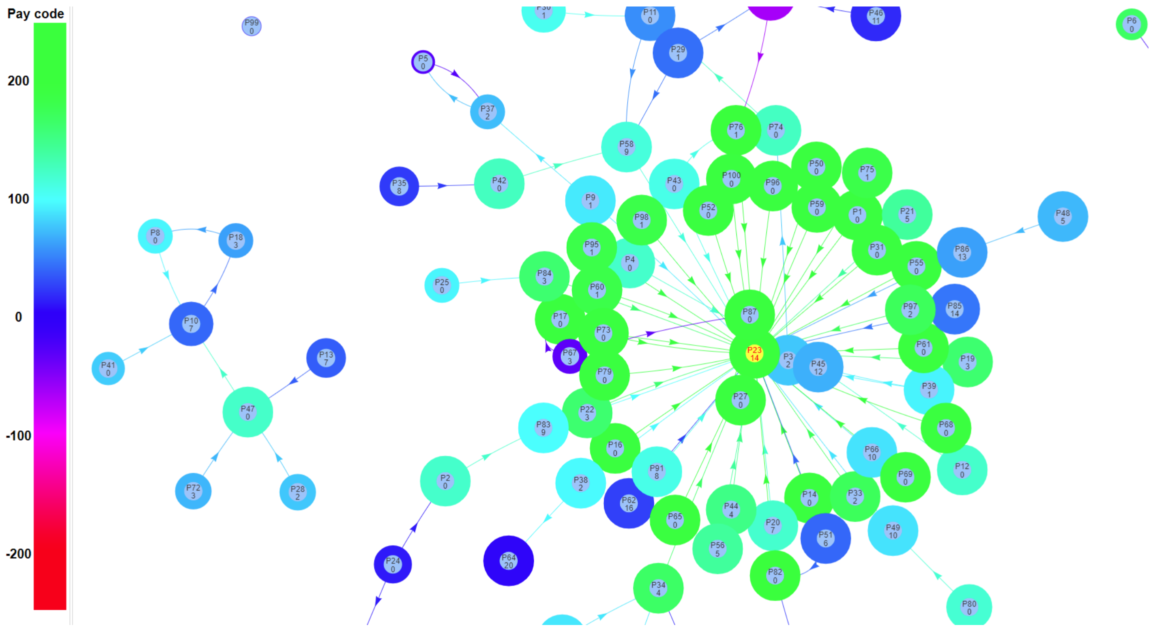
One possible explanation for the high efforts is that individuals enjoy non-monetary benefits from being a hub, and this incentive is reinforced in larger groups. An alternative hypothesis is that in larger groups, the complexity of the dynamics overwhelms individuals and they are led into large efforts, in spite of the lower payoffs. To examine these explanations, we design a treatment in which subjects are shown the payoffs of everyone. The provision of payoff information on everyone facilitates comparison of payoff performances with others. This may make it easier for subjects to understand the payoff implications of their own choices. Availability of such information may also alter individual behavior due to possibilities of imitation or sophisticated learning (Schlag (1998), Huck, Normann, and Oechssler (1999), and Camerer (2003)).

Figures 3 and 4 present snapshots taken in the payoff information treatment with a hundred subjects. Observe that the specialization in linking continues to hold in this setting. However, there is a major change in the behavior of individuals seeking to become a hub: the most connected individual (P23) starts at a high effort 14, but then shades her efforts. The key difference with the baseline is that the outcome is closer to a "pure connector outcome" in some groups and, in most groups, the most connected individual earns much more than the peripheral individuals.

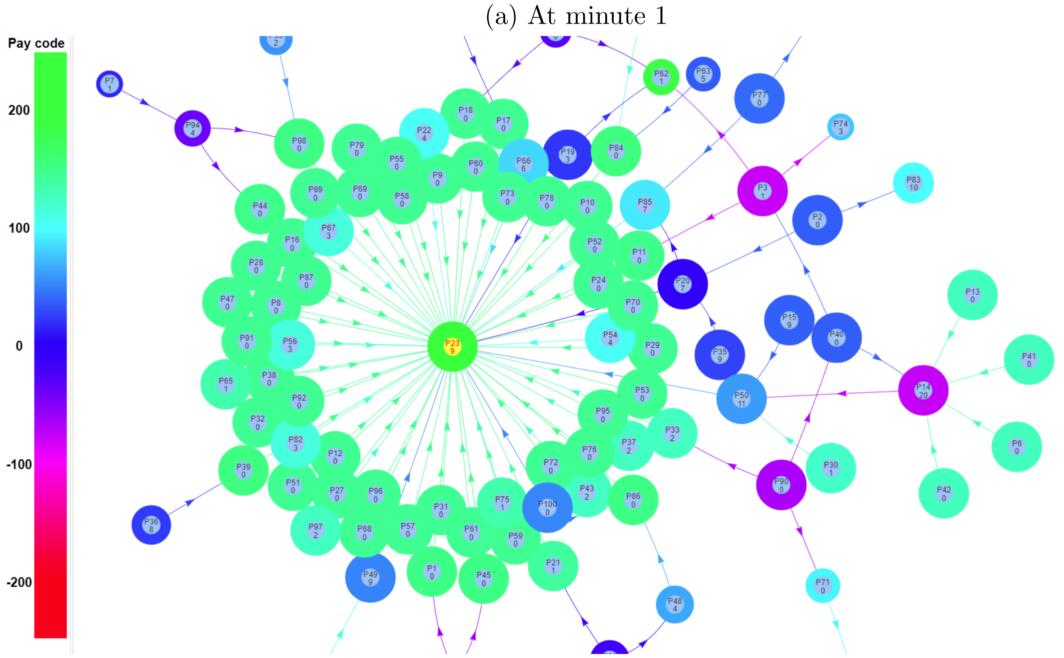
For the payoff information treatment we find that, in all four group sizes, there is specialization in linking and in efforts, and this specialization is more transparent in the large groups. We also find that in the payoff information treatment in small groups there is a strong positive correlation while in large groups there is a weak correlation between connectedness and effort. Indeed, in some large groups the most connected individual puts in 0 effort leading to the pure connector outcome (as in Figures 3 and 4). The pure connector outcome is in sharp contrast to the pure influencer outcome observed in the baseline treatment.

These powerful treatment effects motivate an examination of individual decision rules. We study the behavior of three types of subjects: most connected, 2nd most connected, and the others. The effort dynamics bring out two broad patterns: one, they show large effects of group size and payoff information on the behavior of the two most connected individuals. Two, other—poorly connected—subjects behave similarly across the group sizes and information treatments: they make low effort that is declining over time and somewhat close to myopic best response efforts.

In order to understand the differential effort choices of the two most connected subjects across treatments, we propose a decision rule on activity level of highly connected individ-

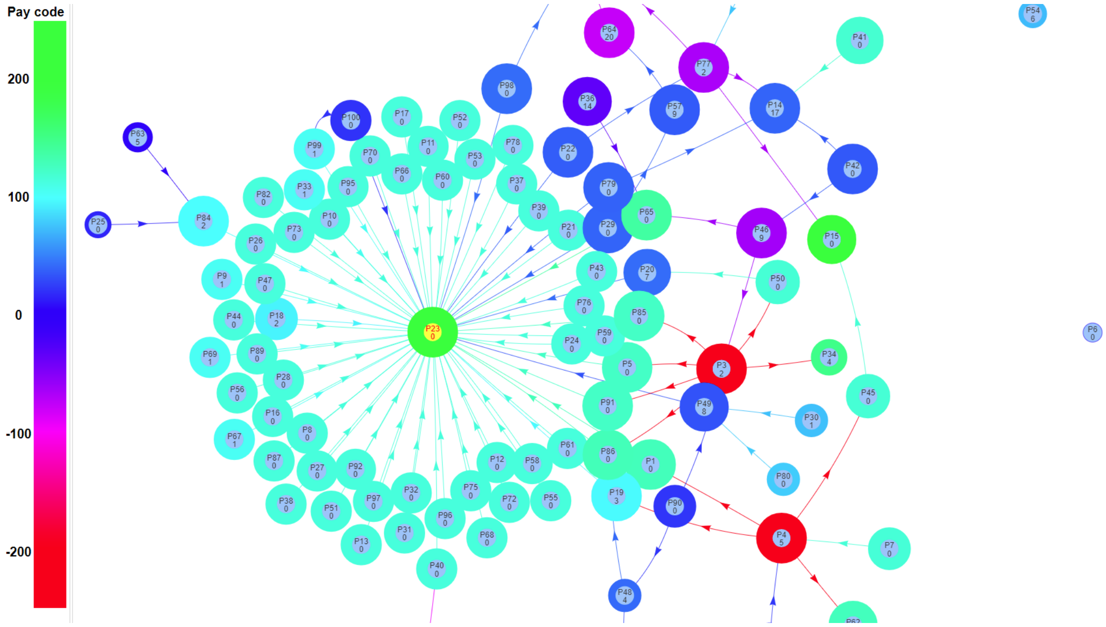


(a) At minute 1

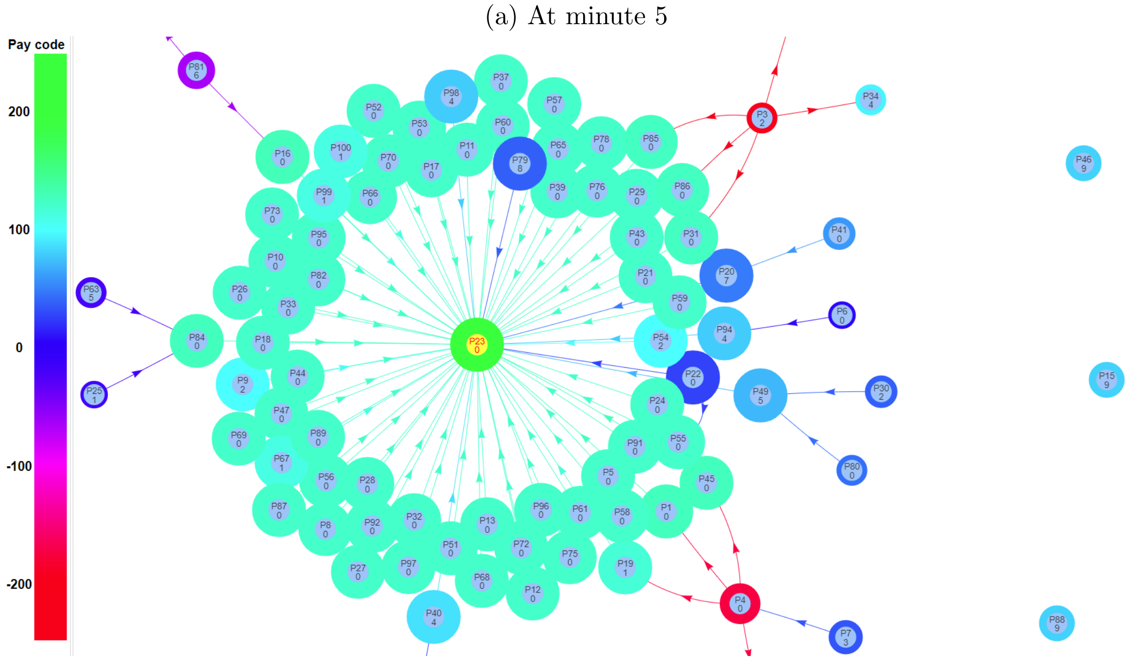


(b) at minute 3

Figure 3: Snap shots with payoff information



(a) At minute 5



(b) At minute 6

Figure 4: Snap shots with payoff information (cont.)

uals that combines myopic best response and competition for hub status. There are two parameters in this decision rule – θ and $\bar{x}$. θ measures the range between higher and lower degree: agents compete if their links are within θ of each other. If they compete then they set effort at $\bar{x}$. So $\bar{x}$ may be interpreted as a measure of their 'aggression'. We estimate θ and $\bar{x}$ that best fit the data in different treatments. There are significant differences in these estimates across treatments. Specifically, we find that $\bar{x}$ is close to the equilibrium level of effort in the pure influencer equilibrium in the small groups (with 4 and 8 subjects) under both information treatments. However, in large groups (with 50 and 100 subjects), there are significant effects of payoff information provision: for instance, for the most connected subject, the estimated value of $\bar{x}$ under payoff information treatment is much lower than it is under baseline information. This suggests that large scale combines with limited information on others' payoffs to make highly connected individuals especially aggressive in their choice of efforts.

Our paper contributes to the study of networks in economics. The principal finding that subjects create networks that are broadly in line with the prediction of the theory on the law of the few. In important early studies, Falk and Kosfeld (2012), Goeree, Riedl, and Ule (2009) have conducted experiments on the pure linking model. In a recent paper van Leeuwen, Offerman, and Schram (2019) test the Galeotti and Goyal (2010) model with linking and activity.[4] A recurring finding in the existing work is that subjects do not abide by the prediction of a hub-spoke network and therefore the model needs to be adapted in different ways to get subjects to create a hub-spoke network. In particular, Goeree et al. (2009) supplement the basic model by creating asymmetries in payoffs across players and van Leeuwen et al. (2019) offer transfers to players based on incoming links. Our finding of specialization in linking and efforts obtains in a baseline model and obtains across a range of treatments – on group size and information provision. These findings therefore constitute a significant departure from the literature. As our findings obtain under parameter settings that are broadly comparable to existing experiments, we attribute the difference to activity being asynchronous and in continuous time.

Our paper contributes to a recent strand of research that has shown the usefulness

[4] We would also like to mention the experimental literature on games in networks (see e.g., Leider, Mobius, Rosenblat, and Do (2009), Charness, Corominas-Bosch, and Frechette (2007), Charness, Feri, Meléndez-Jiménez, and Sutter (2014), Chandrasekhar, Larreguy, and Xandri (2019)) and on games in which players choose partners and then play a coordination game (see e.g., Riedl, Rohde, and Strobel (2016), Kearns, Judd, and Vorobeychik (2012)). Our experiment on the Galeotti and Goyal (2010) model supplements this latter strand of work. The novelty is that actions are asynchronous and in continuous time.

of conducting experiments in continuous time, see e.g., Friedman and Oprea (2012) and Calford and Oprea (2017) and especially the early work of Berninghaus et al. (2006) on network formation. Existing studies are built on an experimental software, called ConG (Pettit, Friedman, Kephart, and Oprea (2014)) and have focused on small group interaction (see e.g., Friedman and Oprea (2012); Calford and Oprea (2017)). The novelty of our paper is that we develop an experimental software that is well suited for the study large group interaction. In order to overcome information overload of evolving networks and relax subjects' cognitive bounds in information processing, our software integrates the network visualization tool with the interactive tool of asynchronous choices in real time. This is achieved by adopting an enhanced communication protocol between the server and subjects' computers. It allows us to run both network visualization and asynchronous dynamic choices in real time without communication congestion and lagged responses, even when participants are interacting remotely from different physical locations.

The paper also contributes to the experimental investigation of group size effects in economic environments (for an overview of the literature, see Ledyard (1994)). In an influential early contribution, Isaac and Walker (1988) show that there is no pure scale effect in contributions in a public good game. On the other hand, Kagel and Levin (1986) present evidence of more aggressive bidding in auctions with common values, giving rise to a larger winner's curse, as the number of bidders grows. Our findings complement this work: subjects seeking to become a hub form more links, choose higher efforts, and earn lower payoffs, as we increase the group size in the baseline treatment. This is in line with a winner's curse result in Kagel and Levin (1986). However, by contrast, in the payoff information treatment, we find that moving from small to large group leads to lower effort and higher earnings for the hub. This finding suggests that information overload is a first order factor that can shape human behaviour in large scale groups.

2 Theory

We present a model of linking and efforts taken from Galeotti and Goyal (2010).

Let $N = \{1, 2, \ldots, n\}$ with $n \geq 3$. Each player $i \in N$ simultaneously and independently chooses a level of effort $x_i \in \mathbf{R}$ and a set of links g_i with others to access their efforts such that $g_i = (g_{i1}, \ldots, g_{ii-1}, g_{ii+1}, \ldots, g_{in})$, and $g_{ij} \in \{0, 1\}$ for any $j \in N \backslash \{i\}$. Let $G_i = \{0, 1\}^{n-1}$. We define the set of strategies of player i as $S_i = \mathbf{R} \times G_i$, and the set of strategies for all players as $S = S_1 \times \ldots \times S_n$. A strategy profile $s = (x, g)$ specifies efforts and the links made by every player. Observe that g is a directed graph; the closure of g is an undirected network denoted by $\bar{g}$ where $\bar{g_{ij}} = \max(g_{ij}, g_{ji})$ for every $i, j \in N$. The undirected link between two players reflects exchange of benefits from efforts. Let $\eta_i(g) = |\{j \in N : g_{ij} = 1\}|$ be the number of links i has formed. For any pair of players i and j in g, the geodesic distance, denoted by $d(i, j; \bar{g})$, is the length of the shortest path between i and j in $\bar{g}$. If no such path exists, the distance is set to infinity. Define $N_i^l(\bar{g}) = \{j \in N : d(i, j; \bar{g}) = l\}$ to be set of players at distance l from i in $\bar{g}$.

Given a strategy profile $s = (x, g)$, the payoffs of player i are:

$$\Pi_i(x, g) = f(x_i + \sum_{l=1}^{n-1} a_l (\sum_{j \in N_i^l(\bar{g})} x_j)) - cx_i - \eta_i(g)k \tag{1}$$

where c denotes the constant marginal cost of efforts, k the cost of linking with another player, and a_l reflects the spillover across players who are at distance l. So if $j \in N_i^l(\bar{g})$, then the value of agent j's effort to i is given by $a_l x_j$. Throughout, it is assumed that $a_1 = 1$, $a_2 \in (0, 1)$, and $a_l = 0$, for all $l \geq 3$. The benefit function $f(y)$ is twice continuously differentiable, increasing, and strictly concave in y. For simplicity, also assume that $f(0) = 0$, $f'(0) > c$, and $\lim_{y \to \infty} f'(y) = m < c$. Under these assumptions, there exists a number $\hat{y} \in X$ such that $f'(\hat{y}) = c$.

There are no general equilibrium characterization results available for this model.[5]. The following result characterizes equilibrium when linking costs are relatively large.

Proposition 1. *Suppose payoffs are given by (1), and that $a_1 = 1$, and $a_2 \in (0, 1)$. Then there exists a $\hat{k}$, such that for $k \in (\hat{k}, c\hat{y})$ the following is true. The equilibrium network is a periphery sponsored star. There exist two possible effort equilibrium configurations:*

- *the pure influencer outcome: the hub invests $\hat{y}$ and everyone else invests 0.*
- *the pure connector outcome: the hub invests 0 and everyone else invests $\hat{y}/(1 + (n - 2)a_2)$.*

Proof. The first step is to observe that in equilibrium, every individual must access at least $\hat{y}$. This is true because if someone is accessing less than $\hat{y}$, then due to the concavity of the

[5] The analysis of Galeotti and Goyal (2010) focuses on polar cases in which $a_1 = 1$ and $a_l = 0$, for all $l \geq 2$ and the case where $a_l = 1$, for all l. Our formulation allows for indirect flow of benefits with decay; this appears to be a natural case.

$f(.)$ function, she can simply increase her utility by raising effort so that the total access equals $\hat{y}$.

The second step is to show that players will form one link or zero link, for sufficiently large linking costs. Observe that an isolated individual will choose $\hat{y}$. So it follows that in a network with connections, no one will ever choose more than $\hat{y}$. Note that if link costs are close to $c\hat{y}$ then it is not profitable to form links with two individuals who each chooses $\hat{y}$. So the only situation in which an individual, A, may choose two or more links arises if an individual accesses significantly more than $\hat{y}$ through each link. Consider a link between A and B. Iterating on optimal effort, it is true that if B chooses $\hat{y}$ then every neighbor of B must choose 0. So A accesses more than $\hat{y}$ only if B chooses strictly less than $\hat{y}$. If a neighbour of B chooses a positive effort, then it must be the case that this person must meet the first order condition on optimal efforts: her total efforts invested and accessed must equal $\hat{y}$. As this person is a neighbour of B, it follows that A cannot access more than $\hat{y}$ via the link with B. So, A will form at most one link in equilibrium.

The third step considers effort configurations. Take the situation in which some individual (say) A chooses $\hat{y}$. It is optimal for everyone else to choose effort 0 and form a link with this person. And it is clearly optimal for A to choose $\hat{y}$ when faced with zero efforts by everyone else.

To conclude the proof, we need to show that the pure connector outcome is the only possible equilibrium in a situation where no player chooses $\hat{y}$. Observe first that the pure connector outcome is an equilibrium so long as $k < c\hat{y}(n-2)a_2/(1+(n-2)a_2)$. Observe that $c\hat{y}(n-2)a_2/(1+(n-2)a_2)$ converges to $c\hat{y}$, as n gets large.

The next step is to rule out any other possible equilibrium. The key observation here is that any equilibrium network must have diameter less than or equal to 2. Suppose the diameter of a component is 3 or more. We know from step 2 that the component must be acyclic. So consider two furthest apart leaf nodes. A variant of the 'switching' argument, developed in Bala and Goyal (2000), shows that one of the two leaf players has a strict incentive to deviate. So every component must have diameter 2. Given that the network is acyclic, this implies it must be a star. It is now possible to apply standard agglomeration arguments to deduce that multiple components cannot be sustained in equilibrium.

Finally, the hub player must choose zero. Suppose not. By hypothesis the hub chooses less than $\hat{y}$. Given that a_1 and $a_2 < 1$, both the hub and the spokes cannot be accessing exactly $\hat{y}$. A contradiction that implies that the hub must choose zero effort.

□

In the pure influencer equilibrium, we witness an extreme version of the 'law of the few': a single person receives all the links formed in society and also carries out all the efforts. The pure connector equilibrium retains the specialization in links: a single person receives all links, but the efforts are evenly spread out. Interestingly, in both equilibria the creation of links is basically egalitarian – $n - 1$ players each form one link. For large k values, the payoff distribution is only slightly unequal in the pure influencer equilibrium. However, the payoff inequality can be very large in the pure connector equilibrium (especially if k is large and a_2 is small). We note that the pure connector equilibrium holds only for a sufficiently large group size n, i.e., $n \geq 2 + k/(a_2(c\hat{y} - k))$.

We now specify the parameters used in the experiment. The function $f(.)$ is taken from Goyal et al. (2017).

$$f(y) = \begin{cases} y(29 - y) & \text{if } y \leq 14 \\ 196 + y & \text{else} \end{cases} \tag{2}$$

For simplicity, the efforts are assumed to take on integer values only and there is an upper bound, $\overline{x} = 20$. So the efforts set is given by $X = [0, 20]$. The cost of effort $c = 11$ and the cost of a link $k = 95$; finally, the decay parameter $a_2 = 1/2$. Given these parameters, it can be checked that $\hat{y} = 9$.

There exists a pure influencer equilibrium in which a single individual chooses 9, all other individuals choose 0 and form a link with the positive effort player. In principle, there exists a pure connector equilibrium in which the periphery players each choose $18/n$, for any $n \geq 50$.[6] Given the integer constraints, this equilibrium is no longer feasible (for $n \geq 50$, $0 < 18/n < 1$ is not an integer). In the treatments with 50 and 100 subjects, the periphery sponsored star where 18 peripheral individuals choose 1 and the rest of the subjects choose 0 constitutes an 'approximate' equilibrium (for details see Online Appendix A).[7] Figure 5 illustrates the pure influencer equilibrium and the pure connector (approximate-)equilibrium.

To summarize, in the pure influencer equilibrium, the hub chooses effort 9, while the spokes choose 0. The hub earns 81, while the spokes each earn 85. In the pure connector equilibrium, the hub chooses effort 0, eighteen spokes choose 1 each, while the other spokes choose 0. The hub earns 198, the active spokes 74, and the inactive spokes 85.

[6] The pure connector equilibrium does not hold in the experimental setting for any $n < 50$.

[7] The periphery player who chooses effort 1 and forms a link with the hub earns 79.25. This person could earn 81 by deleting the link and instead choosing effort level 9.

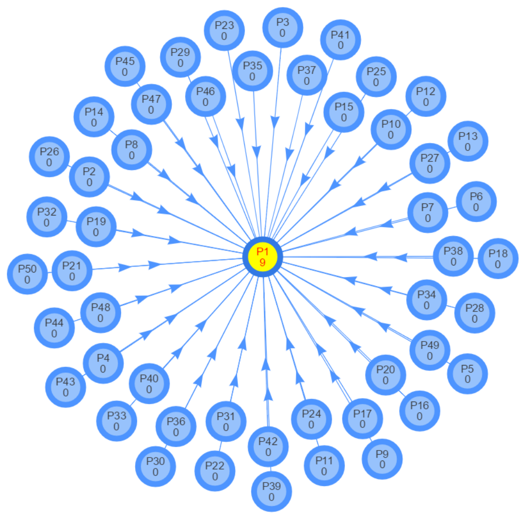


Figure 5: Pure influencer and pure connector equilibrium, $n = 50$

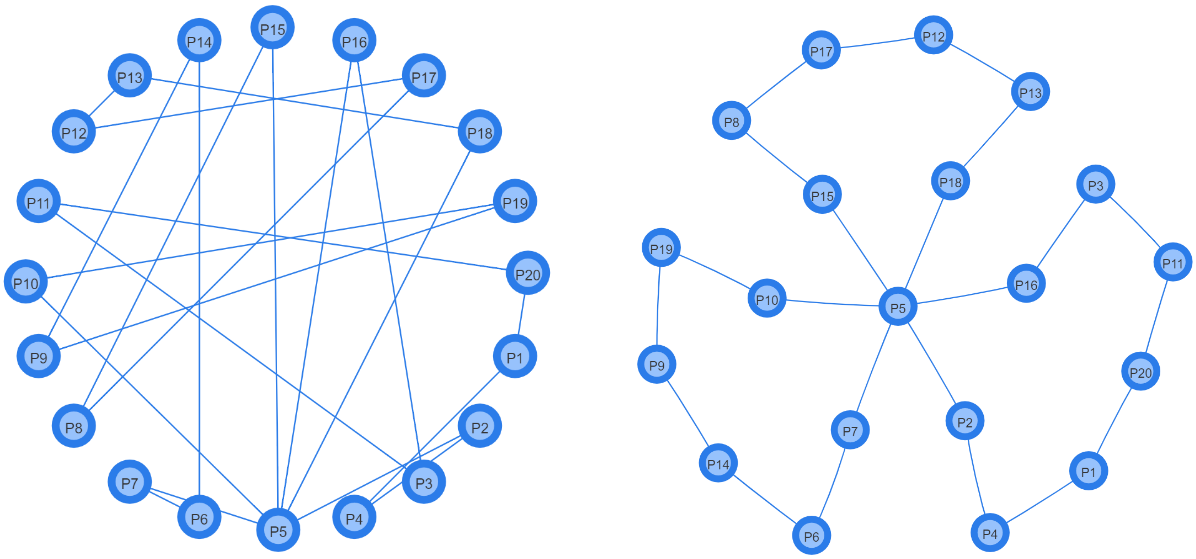
3 Experiment

3.1 Challenges and methodology

As the complexity of subjects' decision making increases in scale, large-scale experiments on network formation pose several major challenges. This section discusses these challenges and explains how our experimental software and design addresses each of them.

Network visualization. Existing studies of network formation in economics have considered small group sizes such as 4 or 8 people and visualized evolving networks with fixed positions of nodes (e.g., Goyal et al. (2017); van Leeuwen et al. (2019)). When the group size increases, such a representation of networks with fixed positions of nodes makes it very difficult for subjects to perceive network features. For example, consider a group of 20 people with fixed positions of nodes in a circle as depicted in Figure 6a; the exact network is barely perceptible by observing this figure.

For subjects to learn their optimal choices, they must have a good idea of the evolving networks. An appropriate tool for visualizing networks is thus critical in running the experiment in continuous time. This leads us to develop an experimental software including an interactive network visualization tool that allows the network to automatically reshape itself based on the evolving structure. This leads us to develop an experimental software



(a) Fixed visualization (b) Adaptive visualization

Figure 6: Examples of network visualization

including an interactive network visualization tool that allows the network to automatically reshape itself in response to decisions made by subjects. We use force-directed algorithms to visualize networks in real time (see, e.g., Eades (1984), Fruchterman and Reingold (1991), Hu (2005), Bostock et al. (2011), and Jacomy et al. (2014)).[8]

Clearly, different network environments will offer differ levels of transparency and information on network architecture. Our strategy is to start with a visualization approach that is efficient and that allows us to systematically explore the effects of different variables – such as scale and variations in information on networks. Of course, the experimental platform is flexible enough to incorporate other ways of representing networks and can be used to explore the effects of network visualization itself on human behavior and network formation. Thus, the experiments reported in this paper could be interpreted as benchmark findings with efficient network visualization.

The network structure in Figure 6a can be represented in a transparent manner in Figure 6b with the network visualization tool we use. In our large-scale experiment, this visualization tool improves graphical clarity of evolving networks and helps subjects dis-

[8]The technical details of the algorithms are provided in Online Appendix B.1 and at the following website: `http://networks.econ.cam.ac.uk/net_formation/experiments.html`.

tinguish between those who are more connected and those who are less connected.

Continuous time with asynchronous choices. It is important to offer subjects adequate opportunities to learn about the environment of decision making, other subjects' behaviors, and how to respond optimally to them. In view of the strategic complexity alluded to above, the issues of learning and behavioral convergence are particularly complicated. To address them, we build on the work of Berninghaus et al. (2006), Friedman and Oprea (2012) and Goyal et al. (2017), and run the experiment in continuous time with near real time updating—of all actions and linking by everyone.[9] At any instant of the game, every subject is free to asynchronously adjust their actions of efforts and linking.

Network information. In addition to the issue of network visualization, there is the issue of network information available to individual subjects. Our platform is flexible with respect to the level of information that is provided on the network. To illustrate the varying degree of network information, consider two extreme scenarios: one, subjects only observe their own neighbors in the current network, and two, subjects get to see the entire network. The information and cognitive load implied by the latter scenario grows rapidly in size of the group. In view of this potential trade-off between transparency of network change and information and cognitive overload as well as the payoff structure of the game, we choose to inform each subject of a local structure of the network within a (geodesic) distance 3.

So given a fixed network, for every subject, we can partition the entire group of subjects into two mutually exclusive subgroups: those who are located within distance 3 from the subject, and those who are located outside this set. Figure 7 provides an illustration of network visualization and information in the experiment with 50 subjects. The left side of Figure 7 shows the group of subjects within distance 3 (and all their links with other subjects within distance 3). The right side of Figure 7 collects the subjects who lie at a distance greater than 3. Observe that in addition to local network information, subjects are informed about every subject's effort—presented as a number within the corresponding node along with that subject's ID. A node's total access to public goods is captured by the size of that node.

Information on Payoffs. We now turn to information on payoffs: clearly subjects need to be able to see their own payoffs in order to learn the profitability of different linking

[9]Although the experimental software allows for real time updating of actions, we voluntarily introduce some latency in our experiment to avoid any possible confusion caused by some overload of activity on the subjects' screen. More precisely, the network depicted on any subject's screen is updated every 5 seconds or whenever the subject makes a decision.

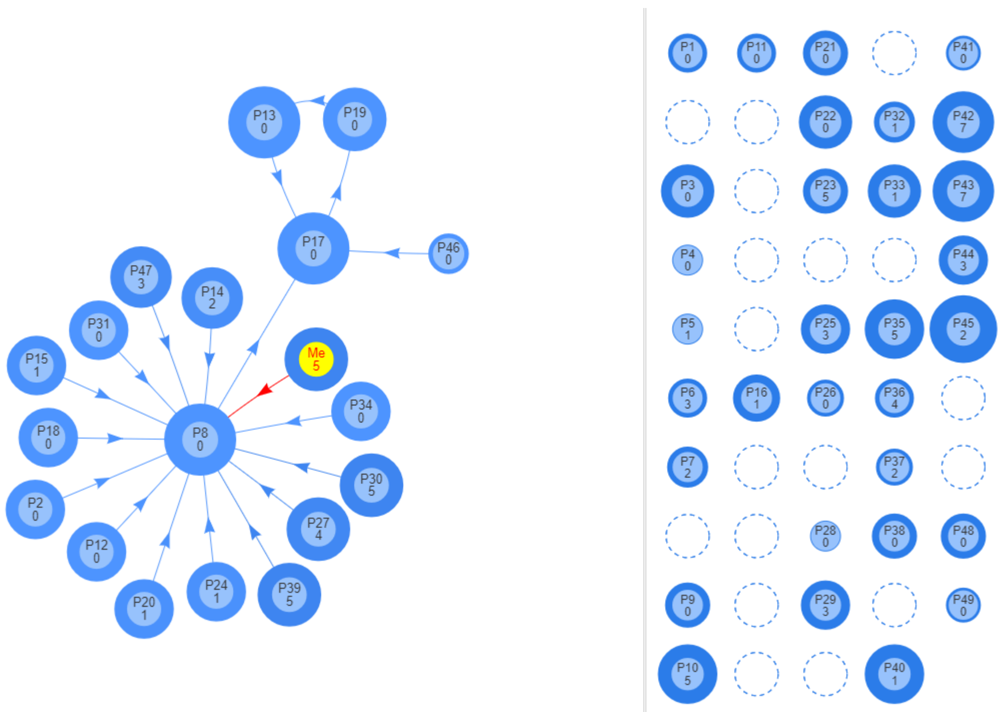


Figure 7: Network information

and effort combinations.[10] What about the information on others' payoffs?

The literature of learning in games provides some guidance on this question, see Camerer (2003) for a survey. In adaptive models such as reinforcement learning and experience-weighted attraction learning (Camerer and Ho (1999)), players ignore information on payoffs of other individuals. In models of imitation learning (Schlag (1998)) and sophisticated learning (Camerer et al. (2002)), players would behave differently if the payoffs of others are known. In the recent body of network experiments (e.g., Goeree et al. (2009) and Falk and Kosfeld (2012)), researchers have tended not to show subjects the payoffs of others. However, when information on others' payoffs is available in particular in large groups where it is difficult to infer such information, subjects may follow a different behavioral rule. In fact, the experimental literature documents that human subjects may behave differently when information on the payoffs of other individuals is available (e.g., Huck et al. (1999)).

Building on these strands of research, it is possible to argue that in games with small groups of subjects, showing the payoffs of others may not be a first order issue, as subjects can compute these payoffs themselves in a fairly straightforward manner. However, in a dynamic game with a hundred subjects—and with the network and efforts configuration constantly evolving—an individual may find it much harder to compute the payoffs of other subjects. The knowledge of others' payoffs may be an important factor in experimental design. The first reason is learning dynamics: observing the others' payoffs could assist subjects in better appreciating the trade-offs associated with different courses of action. The second reason is fairness considerations: the two equilibria described in Proposition 1 exhibit very different level of payoff inequality across players. The pure-influencer equilibrium exhibits a minor payoff difference between the hub player and the spoke players, whereas the pure-connector equilibrium yields a much larger payoff difference between the hub player and the spokes players. These considerations motivate treatments in which we vary the level of information on others' payoff.

In the baseline treatments, subjects are shown their own payoffs but *not* others' payoffs. A subject is also shown the efforts and public good access for all other subjects, as shown in Figure 7. In principle, therefore, a subject can infer the gross payoffs of any subject. But we believe that such inference would be challenging for subjects during a large scale continuous-time game, where the network and effort levels are evolving rapidly. To facilitate

[10] Details about the costs and benefits are provided to the subjects to facilitate their comprehension of their own payoff, as illustrated in Online Appendix D

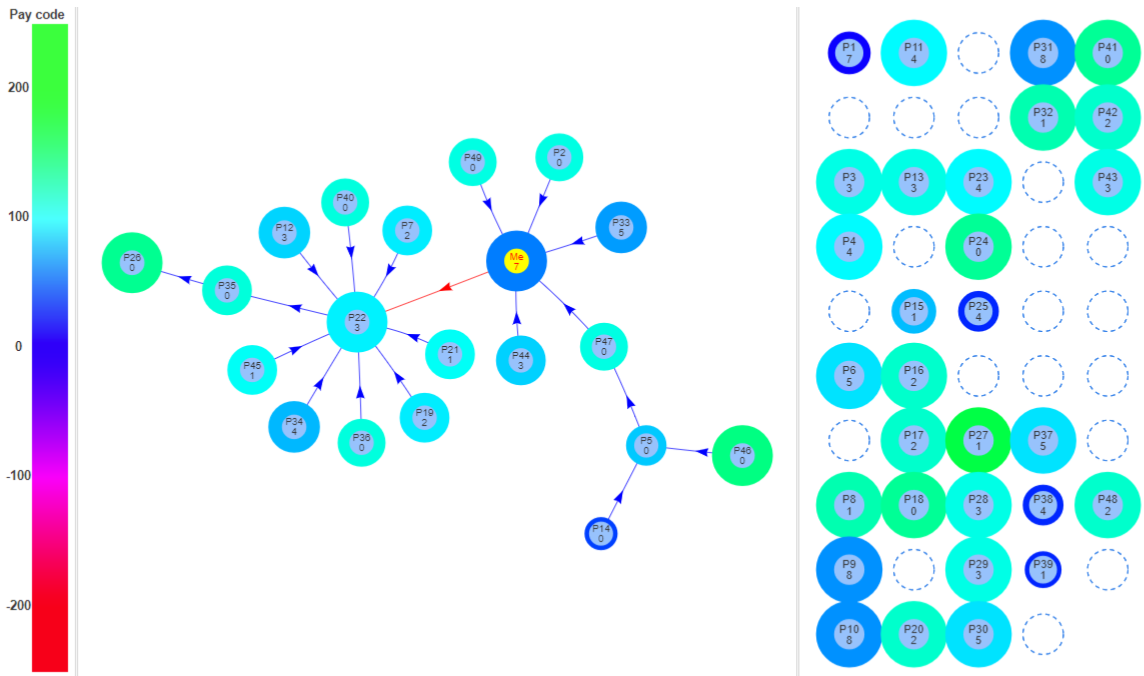


Figure 8: Screen shot of the Payoff Information Treatment

learning, we add information about every player's payoff through a set of color codes as illustrated by Figure 8. Specifically, the border of every node is coloured: the colour varies from green (high positive payoff) to red (high negative payoff). The scale of the colour code is presented at all times on the left hand side, as in Figure 8.

3.2 Treatments and design details

We vary the group size $N \in \{4, 8, 50, 100\}$ and the visibility of others' payoff. Table 1 summarizes the 4×2 structure of our experiment.

All the treatments are based on the payoff function (1). Recall that the marginal cost of effort is set to $c = 11$. This implies an optimal effort of $\hat{y} = 9$. The cost of linking $k = 95$. We restrict effort x as any positive integer value not exceeding 20, i.e., $\overline{x} = 20$. Finally, we set $a_1 = 1$, $a_2 = 0.5$, and $a_l = 0$, for all $l \geq 3$.

At any instant in the 6 minutes game, a subject is free to asynchronously adjust their actions of efforts and linking. For the linking choice, the subject can form or delete a

<table>
<tr><td colspan="2"></td><td colspan="4">Group size</td></tr>
<tr><td colspan="2"></td><td>$N = 4$</td><td>$N = 8$</td><td>$N = 50$</td><td>$N = 100$</td></tr>
<tr><td rowspan="2">Others' payoff information</td><td>NO</td><td>Baseline4</td><td>Baseline8</td><td>Baseline50</td><td>Baseline100</td></tr>
<tr><td>YES</td><td>PayInfo4</td><td>PayInfo8</td><td>PayInfo50</td><td>PayInfo100</td></tr>
</table>

Table 1: Experimental Treatments

link with any other subject by simply double-clicking on the corresponding node in the computer screen. If the subject forms a link with another subject on the right side of the screen (i.e., someone who is in more than 3 geodesic distance away), that subject along with his neighbors and neighbors' neighbors would be transferred to the left side of the computer screen. In a case where the subject removes a link with another subject on the left side of the screen, that subject would be transferred to the right side of the computer screen if they go more than 3 links apart and would remain in the left side of the screen otherwise.

During the experiment, each subject can also choose any level of effort by moving a slider varying from 0 to 20 by increments of 1. This slider is provided on top of the decision screen along with other payoff-relevant information including the subject's gross earnings (i.e., the benefit $f(x)$ where x is the total amount of information the subject has access to), cost of effort, cost of linking, and resulting earnings (i.e., payoff $\Pi_i(x_g)$). Further information on the screen is provided in Online Appendix D.

3.3 Experimental procedures

The experiment was conducted at the Laboratory for Research in Experimental and Behavioral Economics (LINEEX) located in University of Valencia and at the Laboratory for Experimental Economics (LEE) that is located at the University Jaume I of Castelln . All the treatments except for $N = 100$ treatments were conducted at the LINEEX. The experimental sessions with $N = 100$ subjects were conducted through an internet connection between LINEEX and LEE (the number of subjects was then evenly distributed across the two locations). Subjects in the experiment were recruited from online recruitment systems of the two laboratories. A subject participated in only one of the experimental sessions. After subjects read the instructions, the instructions were read aloud by an experimenter to guarantee that they all received the same information. While reading the instructions, the subjects were provided with a step by step interactive tutorial which allowed them to get familiarized with the experimental software and the game. Subjects interacted through a web browser (Google Chrome) on computer terminals and the experimental software was programmed using HTML, PHP, Javascript, and SQL. Sample instructions and interactive tutorials are available in Online Appendix C.

There were in total 18 sessions: 1 session of 16 subjects for each of the Baseline4 and PayInfo4 treatments, 1 session of 32 subjects for each of the Baseline8 and PayInfo8 treatments, 4 sessions of 50 subjects for each of the Baseline50 and PayInfo50 treatments, and 3 sessions of 100 subjects for each of the Baseline100 and PayInfo100 treatments. In each experimental session, subjects were (randomly) matched into a fixed group (if there were more than one group in a session) and interacted with the same subjects throughout the experiment. Therefore, there are 4 independent groups for each of the $N = 4$, $N = 8$, and $N = 50$ treatments and 3 independent groups for each of the $N = 100$ treatments. A total of 1096 subjects participated in the experiment.

The experiment consists of 6 rounds of the continuous-time game, each of which lasted for 6 minutes with the first minute as a trial period and the subsequent 5 minutes as the game with payment consequence. At the end of each round every subject was informed, using the same computer screen, of a moment randomly chosen for payment, detailed information on subjects' behavior at the chosen moment including a network structure and all subjects' efforts, and the resulting earning of the subject. While the membership of a group was fixed within a session, subjects' identification numbers were randomly reassigned at the beginning of every round in order to reduce potential reputation effects. The first round was a trial round with no payoff relevance and the subsequent 5 rounds were effective for subjects' earnings. In analyzing the data, we will focus on subjects' behavior and group outcomes from the last 5 rounds. At the beginning of the experiment, each subject was endowed with an initial balance of 500 points and added positive earnings to or subtracted negative earnings from that initial balance. Subjects' total earnings in the experiment amounted to the sum of earnings across the last 5 rounds and the initial endowment. Earnings were calculated in terms of experimental points and then exchanged into euros at the rate of 100 points being equal to 1 euro. Each session lasted on average 90 minutes, and subjects earned on average about 18 euros (including a 5 euros show-up fee).

At the end of the experiment, subjects took incentivized tasks to elicit social preferences and risk preferences. They are a modified version of Andreoni and Miller (2002) and

Holt and Laury (2002), respectively. In addition, subjects answered a brief version of the Big Five personality inventory test adapted from Rammstedt and John (2007), a comprehension test related to the experimental game, and a debriefing questionnaire including demographic information. More details about these facts can be found in Online Appendix E.

3.4 Connecting theory and experiment

The static model focuses on the trade-off between personal efforts and linking with others and reveals that individual incentives and strategic interaction lead to a network has a very specific structure and only one of two possible effort configurations. This sets a clear line for the experiment. Our interest is in understanding network formation in large groups. To facilitate individual experimentation and learning, we consider a design in continuous time with asynchronous choice: this offers ample scope for experimentation and learning. However, this dynamic game opens the possibility of signalling, cheap talk, and reputation building, forces that go far beyond the original static game. The mapping from the static theory to the experiment is therefore not straightforward.

Our philosophy is that if the arguments in the static model are robust, then subjects should abide by the predictions of the theory in an experimental setting that incorporates realistic elements (such as dynamic linking and effort choice) more accurately. Keeping this in mind, for the purposes of the experiment, we take the following 'high level' view of the predictions of the theory:

1. *Law of the Few*: a small fraction of individuals receive most of the links and carry out most of the efforts. An increase in group size leads to greater specialization in linking and efforts.

2. *Strategic uncertainty*: there exist multiple equilibria; these equilibria differ in actions, linking and payoffs across individuals.

As the costs of effort are linear and there is distance based decay, for any given level of effort, the hub-spoke network maximizes aggregate player welfare. Thus the star is the efficient network architecture. Our computations on payoffs in pure influencer and pure connector equilibrium also reveal that the former is more equitable.

When we combine our high level view of theory with themes in the experimental literature on the role of efficiency (Charness and Rabin (2002)) and inequality aversion (Fehr and

Schmidt (1999), Bolton and Ockenfels (2000)) we are led to the hypothesis that subjects should abide by the pure influencer equilibrium for all group sizes.

This leads us to propose the following hypotheses.

Hypothesis A *Under both information treatments and across all group sizes, specialization in linking and efforts occurs.*

Hypothesis B *Under both information treatments and across all group sizes, linking are positively related to efforts so that efforts of the most connected subjects are higher than those of the other subjects.*

Hypothesis C *Under both information treatments and across all group sizes, the association between linking and payoffs is negligible with little payoff inequality.*

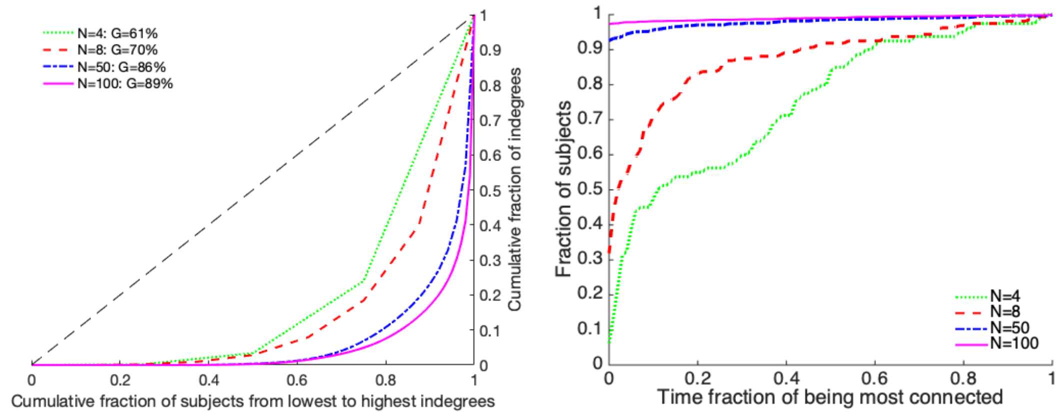
4 Results: Baseline Treatments

We highlighted three key points from Figures 1 and 2: (*i*) extreme specialization in linking and efforts; (*ii*) very large efforts and intense competition among a few subjects to become the hub; and (*iii*) the emergence of the pure influencer outcome. In this section, we examine the experimental data more systematically.

For simplicity in all the data analyses that follow, the data used from every round of the game consists of 360 observations (snapshots of every subject's choices in the group) selected at regular time intervals of one second. Although some information about choice dynamics between two time intervals may be lost, we consider the possible impact of such a simplification as negligible to our analyses. Moreover, unless stated otherwise, all analyses are focused on data from the last 5 minutes of each round of the game.

4.1 Macroscopic Patterns

We start by examining specialization in linking. The first statistic we consider is the number of connections. For any individual, their *indegree* is the number of incoming links from other individuals. The interest is in the specialization/inequality in the indegree. We present two different ways of looking at this issue. The Lorenz curve plots the cumulative fraction of subjects, ranked from least connected to most connected, against the cumulative fraction of total indegrees. The (instantaneous) Lorenz curves are then averaged across seconds of the last five minutes, across rounds, and across groups in each treatment. Figure 9a



(a) Lorenz Curves and Gini Coefficient of Linking

(b) Time fraction

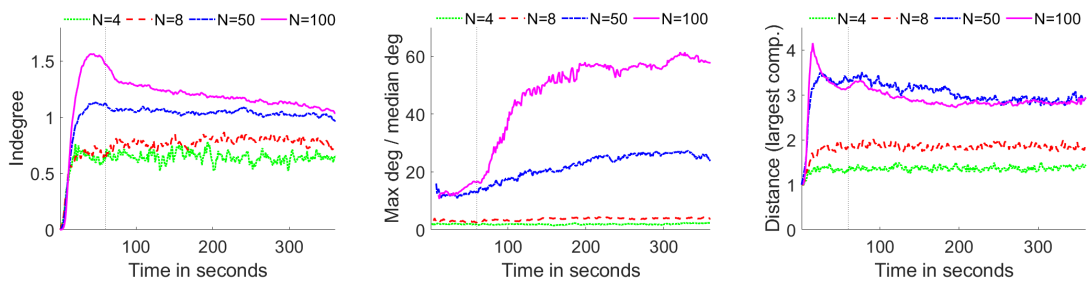
Figure 9: Linking in the baseline treatments

presents these (averaged) Lorenz curves and the corresponding Gini coefficients of indegree across different group sizes. They reveal that specialization is present in every group size, but that it becomes especially striking as the group size increases. This is well reflected in the Gini coefficient measure: it is 0.61 for Baseline4, 0.70 for Baseline8, 0.86 for Baseline50, and 0.89 for Baseline100. By organizing the group-level average data, we conduct t-test for the null hypothesis on the equality of Gini coefficients between a small group ($N = 4$ or $N = 8$) and a large group ($N = 50$ and $N = 100$). We reject it with 5% significance level.

Consider next the number of individuals who become hubs. Specifically, we consider the time fraction (number of seconds out of 5 minutes) for which the individual is most connected. Figure 9b shows the cumulative distributions of time fraction of being most connected and mean indegree ratio. The fraction of subjects who *never* become the most connected player are very high for the large group treatments—0.97 for Baseline100 and 0.93 for Baseline50; this fraction is significantly lower for the smaller groups—0.31 for Baseline8 and 0.06 for Baseline4. It suggests that only a few subjects had any chance of being most connected in the large group treatments, whereas most of the subjects in the small group treatments experienced moments when they were most connected. In this sense, specialization grows with scale.

In addition, we investigate three other properties of network structure – (i) network sparsity, (ii) inequality of linking, and (iii) network closeness. We use average per capita indegree as a measure of sparseness of a network, the ratio of highest degree divided by the median degree as a measure of inequality of linking, and average distance between two nodes as a measure of network closeness.

Figure 10 summarizes our findings about network structure. Firstly, subjects create sparse graphs across all group sizes in the baseline treatments. Average indegree is less than 1 in the small groups. In the baseline50 treatment, the average indegree is stable around 1 over time. In the Baseline100 treatment, average indegree is falling over time to reach 1 at the end of the game. Recall that in the star network, the average indegree would be (roughly) equal to 1. Secondly, linking inequality increases in group size: the ratio reaches around 2 in $N = 4$, 4 in $N = 8$, 27 in $N = 50$, and 60 in $N = 100$ (two-sample t-test with the group level average data: $p < 0.01$ for any pair of group sizes). Thirdly, average distance in the largest components of the networks is no larger than 2 in the small groups, while it converges to 3 in the large groups.[11] Recall that the average distance in a star network would be close to 2. Therefore, we conclude that subjects create networks that are sparse, unequal, and have small average distance across all group sizes in the baseline treatment.

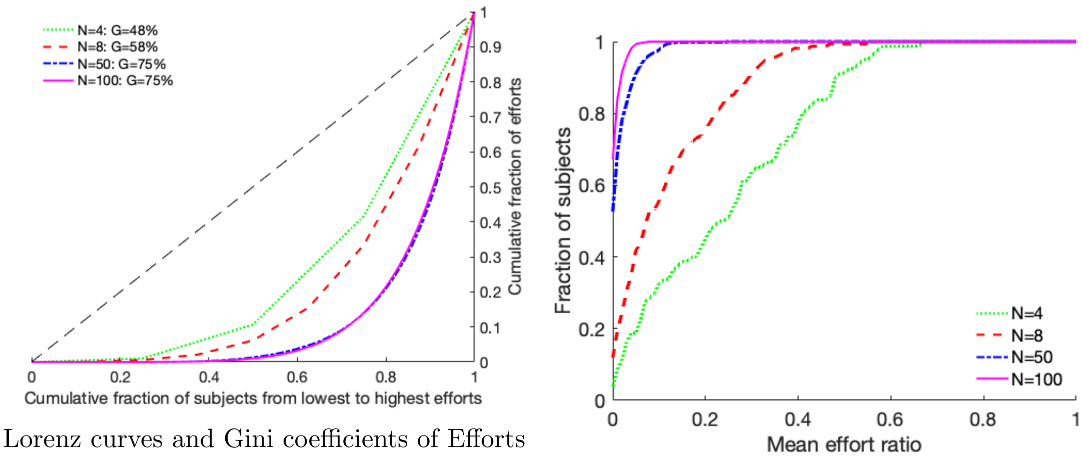


(a) Indegree (b) Link inequality (c) Distance

Figure 10: Network Structure in the baseline treatments

We turn to efforts. Again, the main indicator is the Lorenz curve and Gini-coefficient. Figure 11a presents (averaged) Lorenz curves and Gini coefficients, across different group sizes. Specialization in efforts is present in every group size and it is especially striking in larger groups. This is well reflected in the Gini coefficient of efforts: 0.48 for Baseline4, 0.58 for Baseline8, 0.75 for Baseline50, and 0.75 for Baseline100. The difference between Gini coefficient in the small group treatment ($N = 4$ and $N = 8$) and that in the large group treatment is statistically significant (p-values < 0.01 from t-test with the group-level

[11]The average size of the largest component is close to the group size in each treatment: 3.3 for Baseline4, 6.4 for Baseline8, 44.9 for Baseline50, and 94.8 for Baseline100.



(a) Lorenz curves and Gini coefficients of Efforts

Figure 11: Distribution of Efforts

data).

In order to look further into the details of specialization, consider a variable of mean effort ratio. An individual's effort ratio at every second is defined as her effort divided by the sum of efforts across individuals at that second. We compute the mean of effort ratios across the five minutes for each individual. With this variable, we compute its cumulative distribution in any round and consider the average across rounds and groups. Specialization in efforts becomes substantially more pronounced in large groups. The fraction of subjects whose mean effort ratio is low increases significantly in group size. For instance, relative frequencies of subjects with mean effort ratio being less than or equal to 0.05 are 0.19 for Baseline4, 0.42 for Baseline8, 0.91 for Baseline50, and 0.99 for Baseline100. The distribution of mean effort ratio for a small group treatment first order dominates that for a large group treatment at the usual significance level (p-value < 0.01 from the Kolmogorov-Smirnov test).

In summary, the patterns of linking and effort provision in the baseline treatments confirm Hypothesis A.

Result 1 *Specialization in linking and efforts is present in all group sizes and becomes significantly higher as group size increases.*

We consider the relation between indegrees, efforts and payoffs. Recall from Proposition 1, that there are two equilibria, corresponding to the pure influencer and the pure connector

Table 2: Regression analysis in the baseline treatments

	Indegree ratio (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	4.47***	4.48***		
	(0.43)	(0.43)		
Effort × Large group	0.62***	0.62***		
	(0.07)	(0.07)		
Indegree ratio (%) × Small group			-0.05	-0.03
			(0.14)	(0.16)
Indegree ratio (%) × Large group			-2.79***	-2.75***
			(0.61)	(0.51)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.580	0.581	0.109	0.157

Notes: Robust standard errors, clustered by individual subject, are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

outcomes. In the former, there is a positive relationship between efforts and indegrees and little relationship between indegrees and payoffs. By contrast, in the latter equilibrium, indegrees are negatively related to efforts and positively associated with payoffs.

To explore whether these predictions are borne out by the data, we run linear regression analysis of mean indegree ratio on efforts interacted with the dummies for small group ($N = 4$ and $N = 8$) and large group ($N = 50$ and $N = 100$) and a median regression of payoffs on mean indegree ratio interacted with the dummies for small group and large group. Table 2 reports the regression results with and without controlling demographic information, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality. All regressions include a constant, a dummy for large group, and dummies for rounds. Robust standard errors, clustered by individual subject, are reported in parenthesis. We use the median regression analysis to minimize the impact of outliers in payoffs.[12]

[12] In Online Appendix F.1 we report the same regression analysis by replacing mean indegree ratio with time fraction of being most connected. The regression results with both variables are quite similar.

This regression brings out the positive correlation between efforts and indegree: this relation is statistically significant and positive in the baseline treatments of small groups and large groups. The regression coefficient for efforts is smaller in the large group treatments than in the small group treatments. This is partly because the range of effort is wider in the large group treatments while the range of the indegree ratio is similar across the treatments. Next, we note that the association between indegree and payoffs is weak and insignificant in the small group baseline treatments. There is, however, a strong negative and significant correlation between linking and payoff in the large group baseline treatments. A one percent increase in mean indegree ratio is associated with 2.75 decrease in median payoff for the large group baseline treatments. These associations are robust to the inclusion of additional controls. We summarize the discussion on the relation among effort, linking and payoff as follows.

Result 2 *There is a positive correlation between effort and indegrees in all group sizes. The correlation between indegrees and payoffs is insignificant in the small groups and significantly negative in the large groups.*

This result shows that subjects behaviour conforms with Hypothesis B but not with Hypothesis C. We will return to a discussion on the reasons for the break down of Hypothesis C below.

We next turn to the treatment effects on efficiency. At any second of the game, we define efficiency as the ratio of observed average payoffs to the best Nash equilibrium payoff, i.e. the pure influencer equilibrium. Figure 12 plots the time series of efficiency across group sizes in the baseline treatments (the horizontal dashed line represents the average payoffs in the pure influencer outcome). We observe positive group size effects on efficiency – the ratio is around 0.9 for $N = 4$, 0.95 for $N = 8$, above 1 for $N = 50$, and around 1.4 for $N = 100$ although there is some falling off toward the end of the game in the large groups ($p < 0.01$ from paired t-test comparing $N = 4$, $N = 8$ or $N = 50$ with $N = 100$ with the group level average data). As will be shown, the reason why the level of efficiency in the large groups is higher than that of the best Nash equilibrium is that a few subjects make excessive effort investments in order to become hubs and this increases payoffs of all other subjects.

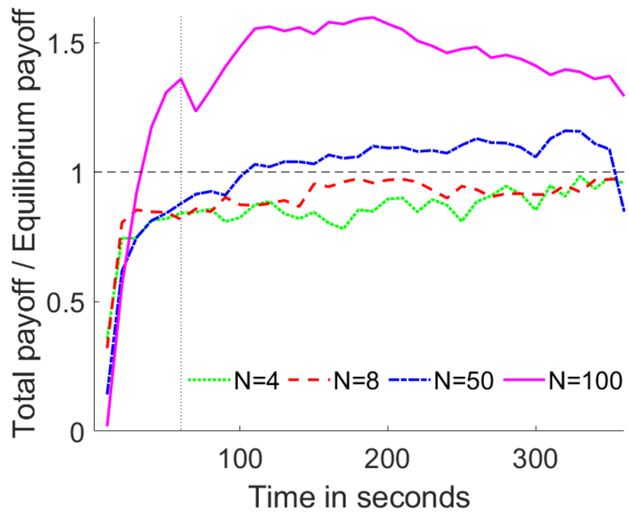


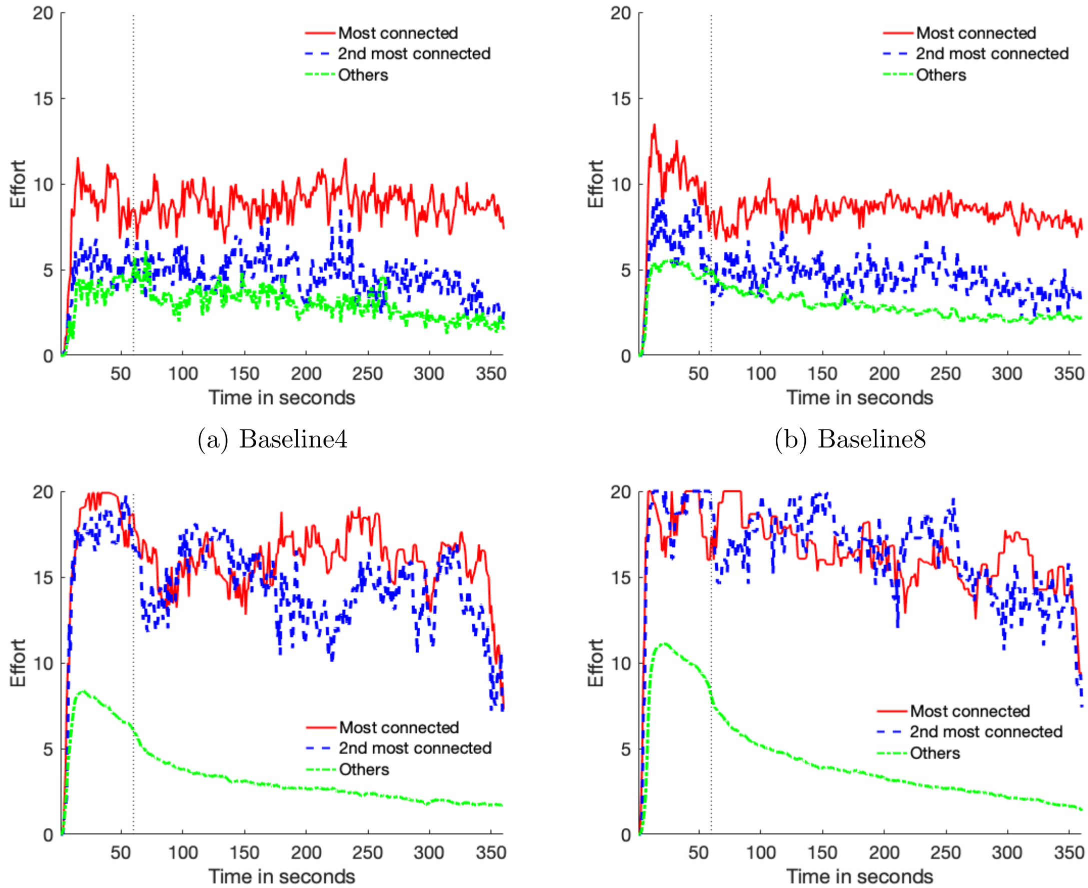
Figure 12: Efficiency in the baseline treatments

4.2 Individual Behavior and Competition Dynamics

This section examines the effects of group size more closely through a study of individual level behavior. The snapshots in Figures 1 and 2 suggest that there are different types of subjects with distinct dynamics of efforts during the game—the two most connected subjects who are competing with each other and the rest of the subjects.

We start with an examination of the dynamics of efforts made by the three different types of subjects identified at every second of the experiment—most connected, 2nd most connected, and the others. Figure 13 presents the average time series of effort for each of the group sizes. The end of the trial minute is represented by the vertical dotted line. We observe very sharp increase in effort by the most connected individual as we move from group size 8 to 50. The other interesting feature of the data is the relative levels of effort between the top two connected individuals: in the small groups there is a persistent gap between their efforts; in the large groups there is a very small gap in effort levels between the top two connected individuals. On the other hand, the average level of effort made by the others is low in all group sizes and steadily decreases over time. These time series patterns suggest that an increase in group size leads to significantly greater competition to become a hub.

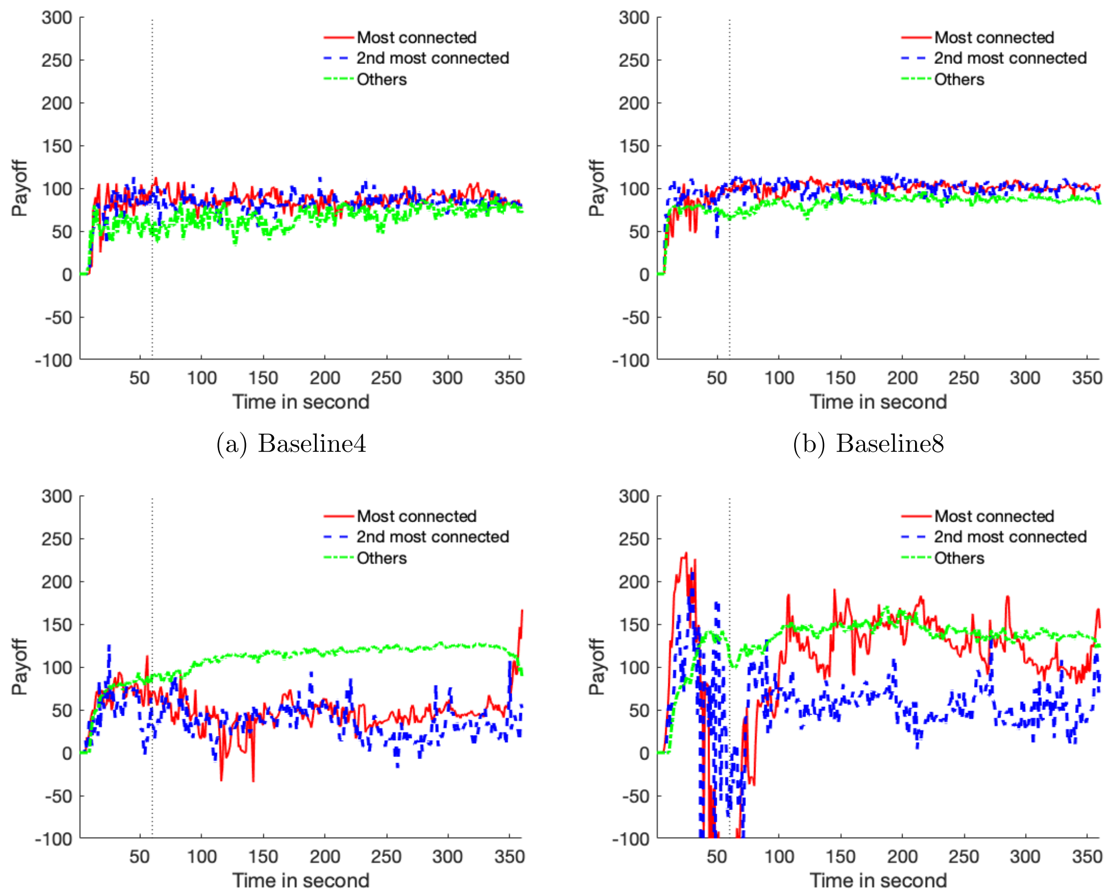
We next look at the dynamics of median payoffs obtained by the three different types



(a) Baseline4 (b) Baseline8

(c) Baseline50 (d) Baseline100

Figure 13: Time series of efforts for the three different types of subjects



(a) Baseline4 (b) Baseline8

(c) Baseline50 (d) Baseline100

Figure 14: Time series of median payoffs for the three different types of subjects

of subjects in Figure 14. The two most connected subjects do not perform better than the other subjects in the large groups. In particular, the 2nd most connected subjects in both Baseline50 and Baseline100 obtain persistently lower payoffs over time than those of the other subjects. This is a consequence of the high efforts. The most connected subjects in the Baseline50 also get persistently lower payoffs than the other subjects except for the last 10 seconds. In the Baseline100, they earn as much as the others for brief periods but the average payoffs are lower than others' payoffs. By contrast, in small groups, the payoffs earned are stable and very similar among the three different types of subjects.

In order to make a statistical assessment on the treatment effects on subjects' behaivor and payoffs, we conduct a linear regression analysis of mean efforts made and a median regression of payoffs obtained by each type of subjects—most connected, 2nd most connected, and the others—on the dummy of large groups ($N = 50$ or 100). In this analysis, we define the types of subjects with the ranking of the fraction of time (across the five minutes) in which a subject is most connected in a round.[13] The most connected individual is the subject who receives the most links for the largest fraction of time. The 2nd most connected individual is similarly defined. We refer to the rest of subjects as the 'others'.

Table 3 reports the regression results after controlling for round dummies, demographic information, comprehension test score, experimental measures of risk aversion and altruism, and personality. Robust standard errors (clustered by individual subject in the regression of efforts) are reported. Average efforts and median payoffs for each type of subjects in the small groups ($N = 4$ and 8) are also reported for comparison.

Table 3 confirms the findings of Figures 13 and 14 and report significant treatment effects on efforts and payoffs. The two most connected subjects made significantly more efforts and earned substantially less in the large groups than in the small groups: 68% more efforts and 27% less payoff for the most connected subject, and 173% more efforts and 55% less payoff for the 2nd most connected subject.[14] Thanks to the intense competition of the two most connected subjects, the other subjects earned 44% more in the large groups than in the small groups.

The discussion on individual behavior, the competition dynamics, and payoffs is summarized as follows.

Result 3 *An increase in group size intensifies the competition between the two most connected subjects. It leads to a significant increase in their efforts and linking . A consequence of this is a decline in their payoffs, relative to the other subjects.*

[13]Figure 25 in Online Appendix F.2 presents histograms showing the time fraction of different efforts over 5 minutes for the three different types of subjects across group sizes in the baseline treatment. The two most connected subjects in the large groups chose the maximum effort level, 20, for the majority of time, whereas they in the small groups chose significantly less with the mode of the most connected subject's effort being around the equilibrium effort level, 9

[14]Tables 7 and 8 in Online Appendix F.1 report the replications of Table 3 by splitting the two large groups. The results remain similar with each of the large groups. In addition, we report the regression analysis of outdegree (the number of links) in Table 9 in Online Appendix F.1. We find that outdegree increases modestly for each type of subjects in the large groups.

Table 3: Scale effects on effort and payoffs in the baseline treatments

	Mean effort			Median payoff		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Large group	6.00***	9.04***	0.62*	-23.75*	-44.94**	37.12***
	(1.05)	(1.10)	(0.32)	(13.25)	(18.13)	(2.90)
Mean or median in small group	8.77	5.24	2.65	86.50	81.00	85.00
Number of observations	75	75	2590	75	75	2590
R-squared	0.59	0.64	0.04	0.19	0.23	0.08

Notes: Robust standard errors (clustered by individual subject in the regression analysis of efforts) are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

5 Payoff Information

We found that, as the group size grows, individuals compete fiercely to become hubs. This leads them to invest very large amounts and, as a result, their earnings suffer. Indeed, in some cases the hubs actually make negative earnings.[15] This is a striking and somewhat unexpected outcome. An obvious concern is that due to the complexity of the environment in large groups some individuals may be keen on competition for hubs without realizing its payoff consequences. If it is indeed the case, individuals may take caution in competing for hubs when their performances are easily comparable to the performances of other people. To examine the scope of this idea, we consider a design in which subjects are shown the payoffs of everyone. This section presents the findings for this treatment.

5.1 Macroscopic Patterns

Figure 15 begins by presenting the average of Lorenz curves and Gini coefficients of indegree across seconds of the last five minutes of the game, across rounds, and across groups in the

[15] We observe that 25% (13%) of the 1st most connected subject's sample in the Baseline50 (Baseline100) earned negative earnings. There is no incidence of negative earnings for the most connected subject in the small group treatments. Negative earnings are often made by excessive efforts and multiple links.

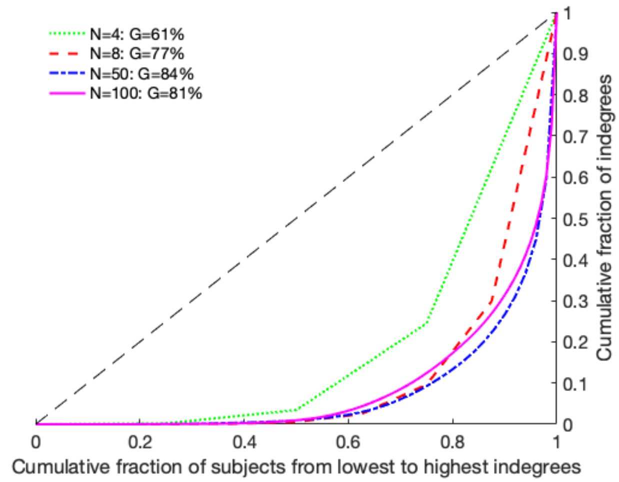


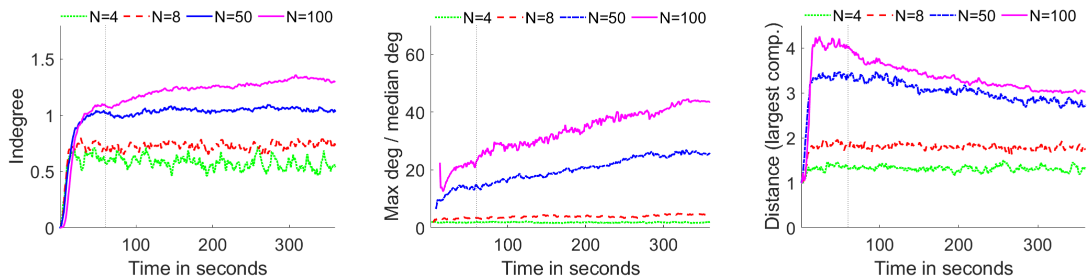
Figure 15: Lorenz curves and Gini coefficients of indegrees: Payoff Information treatments

payoff information treatment. We again observe the aggregate effect of scale on indegree distribution in the payoff information treatment, albeit to a lesser degree compared to the baseline treatments. The Gini coefficients are larger in the large group sizes than in the small group sizes: 61% for PayInfo4, 77% for PayInfo8, 84% for PayInfo50, and 81% for PayInfo100. Comparing these statistics from the baseline treatments, we observe an increase of Gini coefficient in PayInfo8 relative to Baseline8 and a decrease of this in PayInfo100 relative to Baseline100. In spite of this, we observe a statistical difference in this measure of specialization of linking between PayInfo4 and each of PayInfo50 and PayInfo100 (p-value < 0.01 from t-test with the group-level average data). These scale effects are also seen in the cumulative distributions of the time fraction of being most connected and mean indegree ratio (see Online Appendix F.2).

We next examine the three other properties of network structure under payoff information. Figure 16 summarizes our findings about network structure. As in the baseline treatments, in the payoff information treatments we observe sparse networks across all group sizes, the increase of linking inequality in group size, and small average distances in the largest components of the networks[16] Hence, we conclude that subjects create networks that are sparse, unequal, and have small average distance across all group sizes in

[16]The average size of the largest component is close to the group size in each treatment: 3.1 for PayInfo4, 6.1 for PayInfo8, 43.5 for PayInfo50, and 93.5 for PayInfo100.

the payoff information treatment.



(a) Indegree (b) Link inequality (c) Distance

Figure 16: Network Structure in the payoff information treatments

Turning to efforts, Figure 17 presents the (averaged) Lorenz curves and Gini coefficients across groups. The scale effects we see are similar to what we observed for the baseline treatment. The Gini coefficient increases in group size: 38% for PayInfo4, 57% for PayInfo8, 74% for PayInfo50, and 74% for PayInfo100. We observe a statistical difference in this measure of specialization of efforts between PayInfo4 and each of PayInfo50 and PayInfo100 (p-value < 0.01 from t-test with the group-level average data) and between PayInfo8 and PayInfo50 (p-value < 0.04). We also observe similar scale effects with the cumulative distributions of mean effort ratio (see Online Appendix F.2).

We next consider the relation between efforts and linking and the relation between linking and payoff in the payoff information treatments. As in the baseline treatments in Section 4.1, we run linear regression analysis of mean indegree ratio on efforts interacted with the dummies for small groups and large groups and a median regression of (median) payoffs on mean indegree ratio interacted with the same dummies for small groups and large groups. Table 4 reports the regression results with and without controlling the set of additional variables. Robust standard errors, clustered by individual subject, are reported in parenthesis.

Firstly, starting with the regression results about the large group payoff information treatments, we find that the relation between efforts and linking is significantly positive but relatively weak. On the other hand, the relation between linking and payoff in the large group treatments is strongly positive. This is against the prediction of the pure-influence equilibrium summarized in Hypothesis C. Overall, we interpret that showing information on others' payoff makes subjects choose efforts cautiously and leads to the relation between linking and payoff which is predicted by the pure-connector equilibrium.

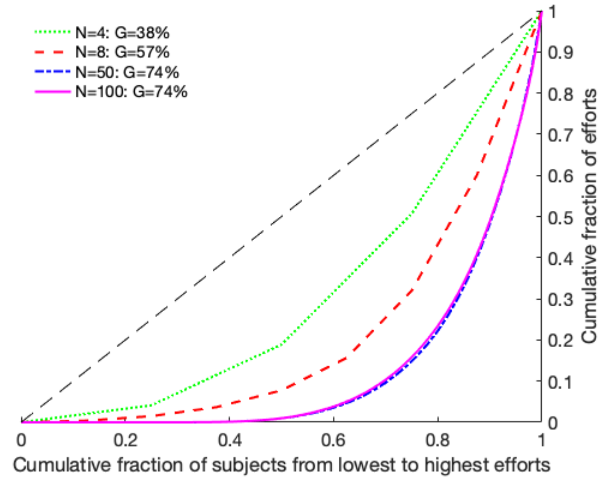


Figure 17: Lorenz curves and Gini coefficients of Efforts: Payoff Information treatment

Table 4: Regression analysis when information on others' payoff is observable

	Indegree ratio (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	5.42***	5.37***		
	(0.55)	(0.55)		
Effort × Large group	0.32***	0.32***		
	(0.06)	(0.06)		
Indegree ratio (%) × Small group			0.26**	0.28***
			(0.12)	(0.11)
Indegree ratio (%) × Large group			1.81***	1.75***
			(0.16)	(0.12)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.521	0.523	0.002	0.015

Notes: Robust standard errors, clustered by individual subject, are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

We next look at the regression results about the small group payoff information treatments. As in the baseline treatments, we find a strong and positive association between efforts and linking in the small group payoff information treatment. This pattern is in line with the corresponding prediction of the pure-influencer equilibrium. When we analyze the relation between linking and payoff, we observe a relatively weak and positive relation between linking and payoff.

These observations are summarized as follows.

Result 4 *(i) Specialization in linking and efforts continues to hold in the payoff information treatments. (ii) In the small groups, the correlation between linking and effort is strongly positive, while the correlation between linking and payoff is weak. (iii) In the large groups, there is a strongly positive correlation between linking and payoffs, while the correlation between efforts and linking is weak.*

A comparison of Results 1-3 with Result 4 reveals that specialization is significant and increasing group size reinforces it under both the baseline and the payoff treatment. Subjects thus conform with Hypothesis A. The second finding is that payoff information interacts powerfully with scale: the negative correlation between degrees and payoffs is reversed. This in turn means that subjects move away from a pure influencer outcome towards a pure connector outcome. We will take up the reasons for this change in behaviour and the breakdown of Hypotheses B and C (in large groups under payoff information treatment) below.

These findings on linking and efforts have implications on efficiency. Figure 18 plots the time series of efficiency across group sizes in the payoff information treatments (the horizontal dashed line represents the average payoffs in the pure influencer outcome). The level of efficiency is slightly below 1 and we do not observe any effects of group size in almost all cases (two-sample t-test with the group level average data: p=0.79 for $N = 4$ and $N = 8$, p=0.05 for $N = 4$ and $N = 50$, p<0.01 for $N = 4$ and $N = 100$, p=0.30 for $N = 8$ and $N = 50$, p=0.97 for $N = 50$ and $N = 100$, and p=0.30 for $N = 8$ and $N = 100$). It is because subjects coordinate on the pure influencer outcome in the small group payoff information treatment, whereas they coordinate on the pure connector outcome in the large groups. Overall, the effects of group size on efficiency interact with whether information on others' payoffs is available.

We now examine individual behavior that gives rise to these scale and payoff information effects.

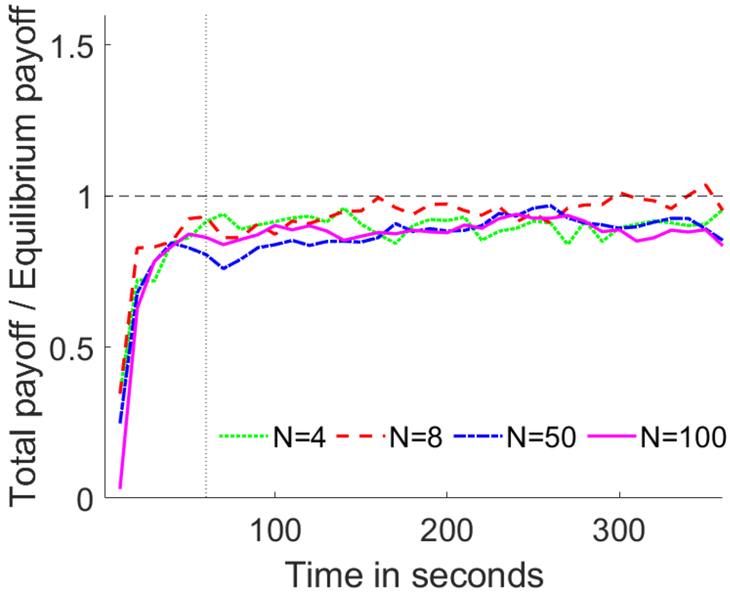
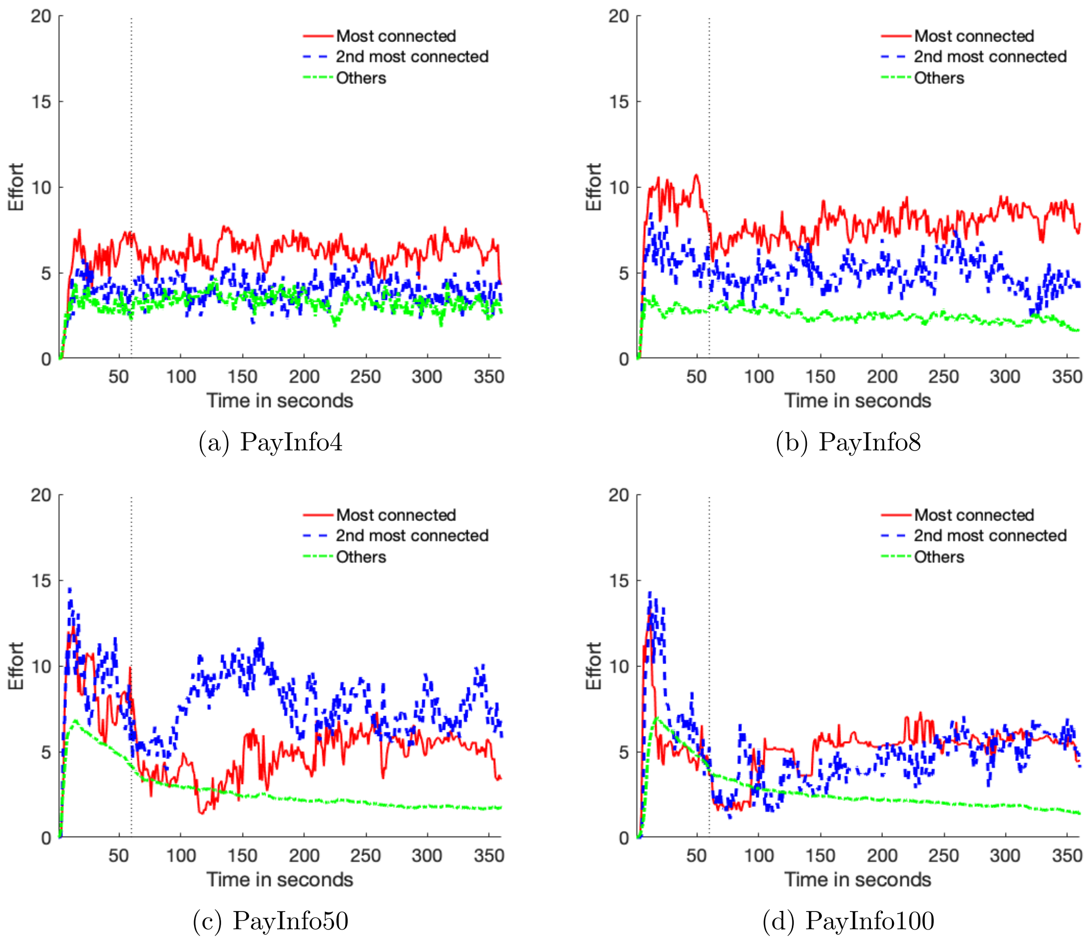


Figure 18: Efficiency in the payoff information treatments

5.2 Individual Behavior and Competition Dynamics

We start with Figure 19 that presents the time series of average efforts (the end of the trial minute is represented by the vertical dotted line) for the three different types of subjects identified at every second in each of the group sizes. Compared to Figure 13 in the baseline treatments, we observe that competition dynamics in the large group payoff information treatments is quite different: the time series of efforts made by two most connected subjects are substantially lower when information on others' payoffs is visible. By contrast, in the small groups, the dynamics of efforts is similar across the payoff information treatment and the baseline. The behavior of 'other' subjects is similar across the two information treatments and across different group sizes.

Figure 20 presents time series of median payoffs for the three types of subjects in the payoff information treatment. We would like to compare this figure with Figure 14 that summarizes the outcomes in the baseline treatments: this comparison reveals that there is a sharp increase in the payoffs of the most connected subjects in the large groups. In the small groups, the payoffs are similar across the two information treatments. Putting together the observations from Figure 19 and Figure 20, we are led to conclude that the availability



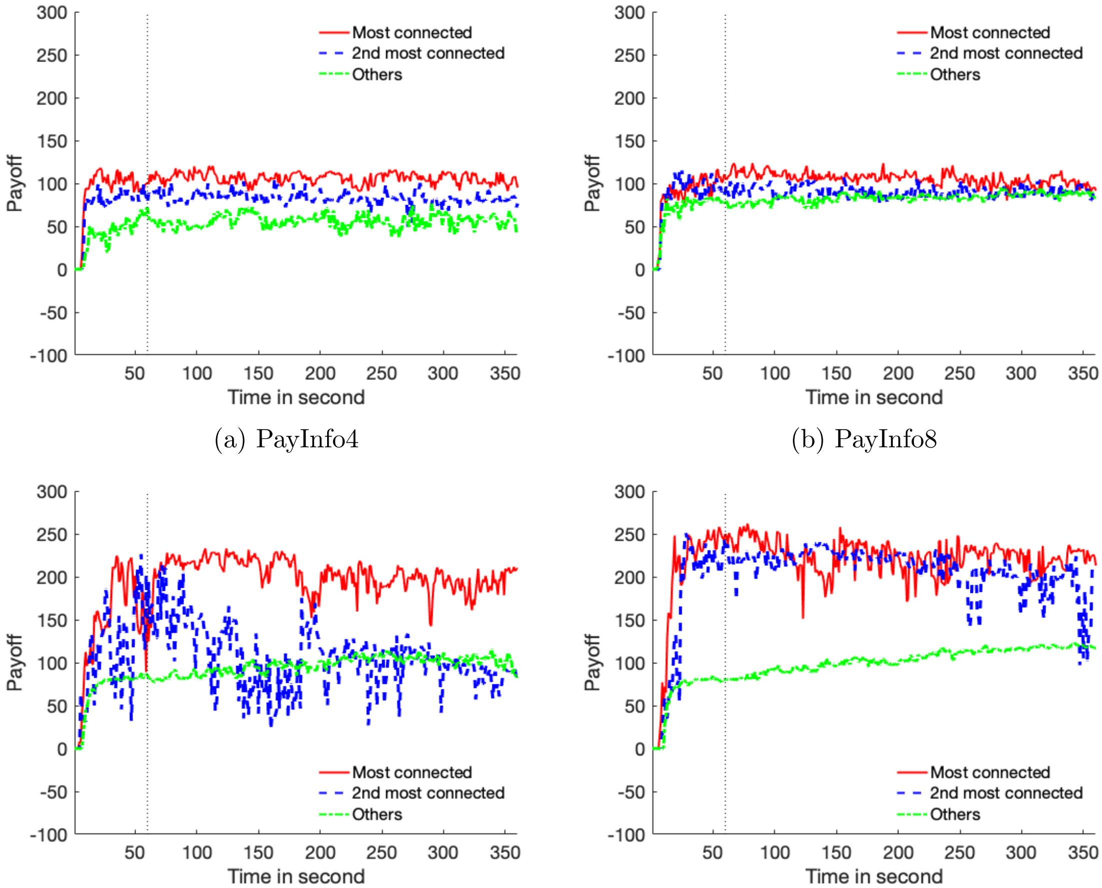
(a) PayInfo4

(b) PayInfo8

(c) PayInfo50

(d) PayInfo100

Figure 19: Time series of efforts in the payoff information treatment



(a) PayInfo4 (b) PayInfo8

(c) PayInfo50 (d) PayInfo100

Figure 20: Time series of median payoffs in the payoff information treatment

of information on others' payoffs leads to lower efforts by the two most connected subjects and this in turn creates large payoff gains for them.

We next conduct a linear regression analysis of mean efforts made and a median regression of payoffs obtained by each type of subjects on the dummies of payoff information and large group ($N = 50$ or 100) and their interaction dummy. As was done in Table 3, we define each type using the time fraction of being most connected in the 5 minutes.[17]

[17]Figure 26 in Online Appendix F presents histograms showing the time fraction of different efforts over 5 minutes for the three different types of subjects across group sizes in the payoff information treatment. We observe a drastic change of efforts in the large groups: in the payoff information treatments, the two most connected subjects substantially lower their level of efforts and the unique mode of the distribution

Table 5: Treatment effects on effort and payoffs

	Mean effort			Median payoff		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Payoff info	-0.75	0.52	0.00	6.71	-12.75***	-10.56***
	(0.77)	(0.70)	(0.36)	(11.54)	(4.53)	(1.97)
Large group	6.30***	8.41***	0.62**	-30.33*	-42.76**	36.20***
	(1.04)	(1.19)	(0.30)	(17.08)	(17.34)	(1.90)
Payoff info × Large group	-9.24***	-9.00***	-0.91**	119.24***	120.76***	-14.07***
	(1.41)	(1.63)	(0.39)	(29.18)	(29.02)	(2.30)
Mean or median in large group baseline	15.12	13.22	3.22	59.00	47.00	126.50
Number of observations	150	150	5180	150	150	5180
R-squared	0.53	0.51	0.05	0.09	0.17	0.09

Notes: Robust standard errors (clustered by individual subject in the regression analysis of efforts) are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 5 reports the regression results controlling for round dummies, demographic information, comprehension test score, experimental measures of risk aversion and altruism, and personality. Average efforts and median payoffs for each type of subjects in the large groups ($N = 50$ or 100) baseline treatment are also reported for comparison. The coefficient of the interaction dummy captures the difference-in-differences effect for the treatments of large group and payoff information. On the other hand, the coefficient of the payoff information describes the payoff information effect in the small groups.

In large groups, we observe significant effects of payoff information on efforts and payoffs. Compared to the corresponding subject type in the large group baseline treatments, the most connected subject makes 61% less effort and earns 202% higher payoffs, and the 2nd most connected subject makes 68% less effort and obtains 260% higher payoffs. The rest of subjects also made 28% less effort and earn 11% lower payoffs in the large group pay-

is at zero effort. On the other hand, we observe little change of efforts in the small group sizes by showing information on others' payoffs.

off information treatments.[18] Hence, the two most connected subjects in the large groups significantly lowered their efforts and gained substantial payoffs when information on others' payoffs is available. In contrast, in the small groups, subjects' efforts did not respond to the availability of information on others' payoff and their payoffs changed modestly.

Result 5 *(i). In the small groups, subjects do not change efforts and linking behavior significantly in response to the availability of information on others' subjects. (ii). In the large groups, the two most connected individuals make substantially lower efforts in the payoff information treatment as compared to the baseline treatment. This results in large payoff gains for them.*

We are led to the view that, when groups are large, it is difficult for a subject to keep track of the payoff implications of different choices. When information on payoffs of others is not directly shown, some subjects are willing to make large efforts in order to attract links from others. In doing so, these subjects appear not to understand that these large efforts lead to much lower payoffs. However, when information on others' payoffs is made directly observable, individuals become more cautious in their effort decisions. This is in contrast with what we observe in small groups: there subjects are able to keep track of the payoff implications of their effort more accurately and the impact of showing information on everyone's payoffs has relatively small impact on subjects' effort behavior.

6 Matching Effort Dynamics with Learning Rules

The effort dynamics presented in Figures 13 and 19 bring out two general points: one, they show large effects of group size and payoff information on the behavior of the two most connected individuals. Two, these figures show that other – poorly connected – subjects behave similarly across the group sizes and information treatments: they make low effort that is declining over time and perform close to myopic best response.[19] By behaving close to myopic best response, a significant majority of subjects of others type earned at least as large as the equilibrium payoff for spokes in the pure influencer equilibrium prediction

[18]Tables 11 and 12 in Online Appendix F.1 replicate Table 5 by considering $N = 50$ and $N = 100$ separately. The negative effect of payoff information remain similar in each case. Regarding outdegree, Table 13 in Online Appendix reports little effect of payoff information on outdegree.

[19]Figures 29 and 30 in the Online Appendix F.2 present the effort dynamics of less connected subjects and compare them to their myopic best response efforts. Myopic best response (with some minor noise) can approximate the observed behavior.

in the second half of the experiment, which is reported in Table 14 in Online Appendix F.1. Consequently, in what follows, we focus on the behavior of the two most connected individuals. We will argue that their effort variations across the group size and payoff information treatments can be accommodated with a decision rule that combines myopic best response and competition for hub status through effort.

Before introducing this decision rule, we first note that the behavior of most connected subjects in the large group baseline treatment is not justifiable from a dynamically rational point of view. A subject can guarantee herself an average payoff of 81 with an effort of 9 and zero links (regardless of what others do). Figure 14 shows that the two most connected subjects reach payoffs lower than this payoff in most cases. Therefore their behavior is not consistent with dynamic optimal choice.[20] This suggests that individuals in the baseline treatment who make large investments do not appreciate that their strategy is not as profitable as other strategies. As these excessive investments happen only in large groups, and it disappears when payoff information for everyone is provided, we are led to the conclusion that increasing group size makes it difficult for highly connected subjects to keep track of the relation between actions and payoffs and this leads to over-investment. Therefore, we propose a boundedly rational decision rule that can account for the behavior of the two most connected subjects across all the treatments.

Let $\{(x_{it}, D_{it})\}_{t=1}^{360}$ denote the sample of individual i's effort (x_{it}) and indegree (D_{it}) over periods, $t = 1, ..., 360$. We use the sample of two most connected individuals i and j who compete to be a hub. Because the computer screen was updated every 5 seconds or whenever the individual made a decision, we allow 3 seconds time lag in defining effort levels predicted by the learning rule.[21] The decision rule we consider has two parameters, $\theta \geq 0$ and $\overline{x} \in [0, 20]$, to be estimated from the data, with the following features:

$$x_{it}\left(\theta_i, \overline{x}_i\right) = \begin{cases} \overline{x}_i & \text{if} \quad |D_{i,t-3} - D_{j,t-3}| \leq \theta_i, \ x_{it-3} > \ x_{jt-3} \\ x_{jt-3} & \text{if} \quad |D_{i,t-3} - D_{j,t-3}| \leq \theta_i, \ x_{it-3} \leq \ x_{jt-3} \\ x_{mbr,t} & \text{if} \quad |D_{i,t-3} - D_{j,t-3}| > \theta_i \end{cases}$$

This decision rule has two features that are worth noting. First, the state of compe-

[20] Figure 31 in Online Appendix F.2 shows the continuation payoff that can be expected by one of the most connected individuals at any moment in time, by averaging the actual payoffs earned from that moment until the end of the game. In the large group baseline treatment, competing for the hub position is not profitable. On the other hand, under the payoff information treatment, it is more profitable in the large group as compared to the small group.

[21] Results presented in this section are robust to different values of time lag around 5 seconds.

tition for hub status is captured by the difference between indegrees obtained by the two individuals. Specifically, if the gap between their indegrees (i.e., $|D_{i,t-3} - D_{j,t-3}|$) is larger than θ, the competition is no longer active, either because an individual is sufficiently ahead of the other or has fallen too far behind. In this situation, the individual is assumed to choose an effort predicted by myopic best response, x_{mbr}. A smaller value of θ represents a tighter range of defining the competition state and thus leads to a lower likelihood of staying in competition. The second part of the rule applies when the two players are in competition to become a hub (i.e., $|D_{i,t-3} - D_{j,t-3}| \leq \theta$): in this situation, the individual chooses either $\overline{x}$ if he chose a higher effort than the other or, otherwise, imitates the other's higher effort. A higher value of $\overline{x}$ implies a more severe competition through effort. Hence, the treatments of group size and payoff information are mediated through $\overline{x}$ and θ to affect the behavior of the two most connected individuals.

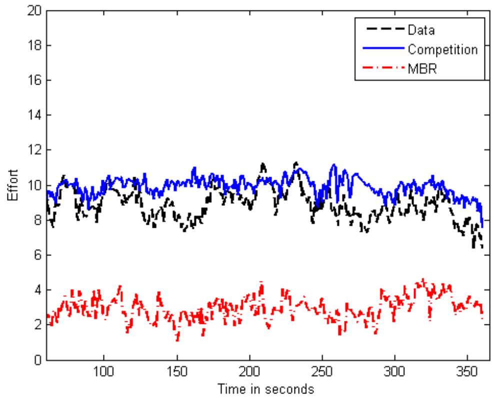
We estimate the two parameters, θ_i and $\overline{x}_i$, for each individual i by minimizing the sum of the distance between observed efforts and those predicted by the decision rule. Table 15 in Online Appendix F.1 presents the estimation results of the parameters across the treatments. First, we observe clear scale effects on the decision rule, in particular on the level of effort made under competition, $\overline{x}_i$. In the small groups under both information treatments, the value of $\overline{x}_i$ is near the effort level of the hub in the pure influencer equilibrium. On the other hand, in the baseline large group treatments, the value of $\overline{x}_i$ is estimated to be much higher. This corroborates our finding of excessive investment in the baseline large group treatments. Secondly, the interaction between payoff information and scale is manifest in the estimated decision rules. In the small groups, the estimated values of the parameters are similar between the two information treatment, suggesting that availability of payoff information in the small groups makes no change in competition for hub status. In contrast, in the large groups, both θ_i and $\overline{x}_i$ appear to be lower in the payoff information treatment as compared to the baseline treatment. For instance, for the most connected subject, the estimated value of $\overline{x}_i$ is 13.3 and 9.3 for PayInfo50 and PayInfo100, whereas it is 18.7 and 16.8 for Baseline50 and Baseline100. These differences of the estimated parameters confirm the interaction of group size and information provision.

In order to assess a goodness of fit of the decision rule, we compare the time series of observed efforts with those predicted by estimated decision rules. Figure 21 and Figure 22 present the fitting of our decision rule to observed effort dynamics of the most connected individual during the last 5 minutes across treatments of group size and payoff information. The black-color dashed line represents effort dynamics observed in the experimental data.

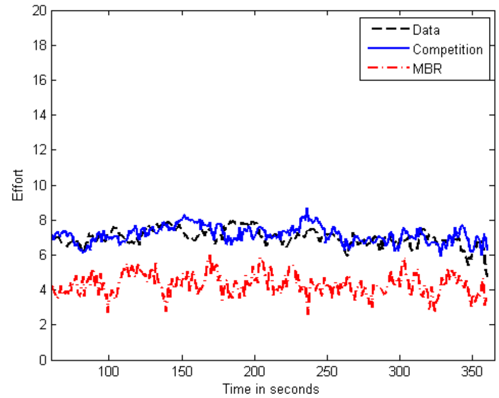
The blue-color solid line describes the dynamics of efforts predicted by the estimated decision rules. For the purpose of comparison, we also draw the time series of efforts predicted by myopic best response (the red-color dotted line). These figures show that our decision rule provides a good fit for the effort dynamics in the experiment; this is specially so if we compare it with the myopic best response rule.

The decision rule applied across the treatments matches the key features of effort dynamics across the treatments. First, in the small groups, the two most connected individuals choose effort close to the effort of the hub in the pure influencer equilibrium from the start of the payoff relevant period till the end of the game. Second, in the large group baseline treatments, the two most connected individuals compete strongly by choosing effort close to the maximum level of effort at the start of the payoff relevant period and decrease their efforts over time. Third, in the large group payoff information treatments, the two individuals tend to start with low efforts, a bit lower than the level of the pure influencer equilibrium and lower their efforts slightly over time. The decision rule we propose captures such patterns of effort dynamics closely. The close fit is obtained through different estimates on the two parameters of the decision rule – θ and $\overline{x}_i$ – as we vary scale and payoff information.

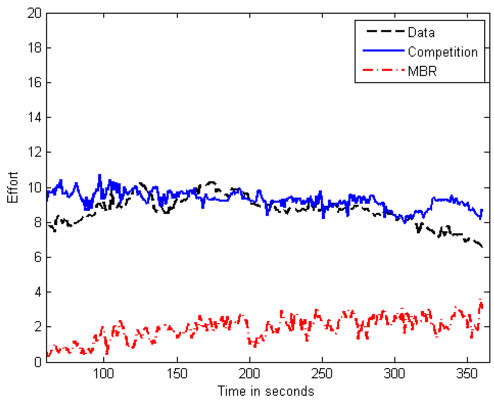
Figure 32 and Figure 33 in Online Appendix F.2 report the same kind of figures for the second most connected individual. The figures show that the decision rule overall performs well in fitting their effort dynamics.



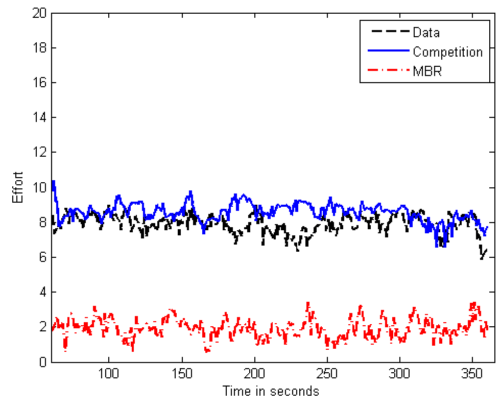
(a) Baseline4



(b) PayInfo4

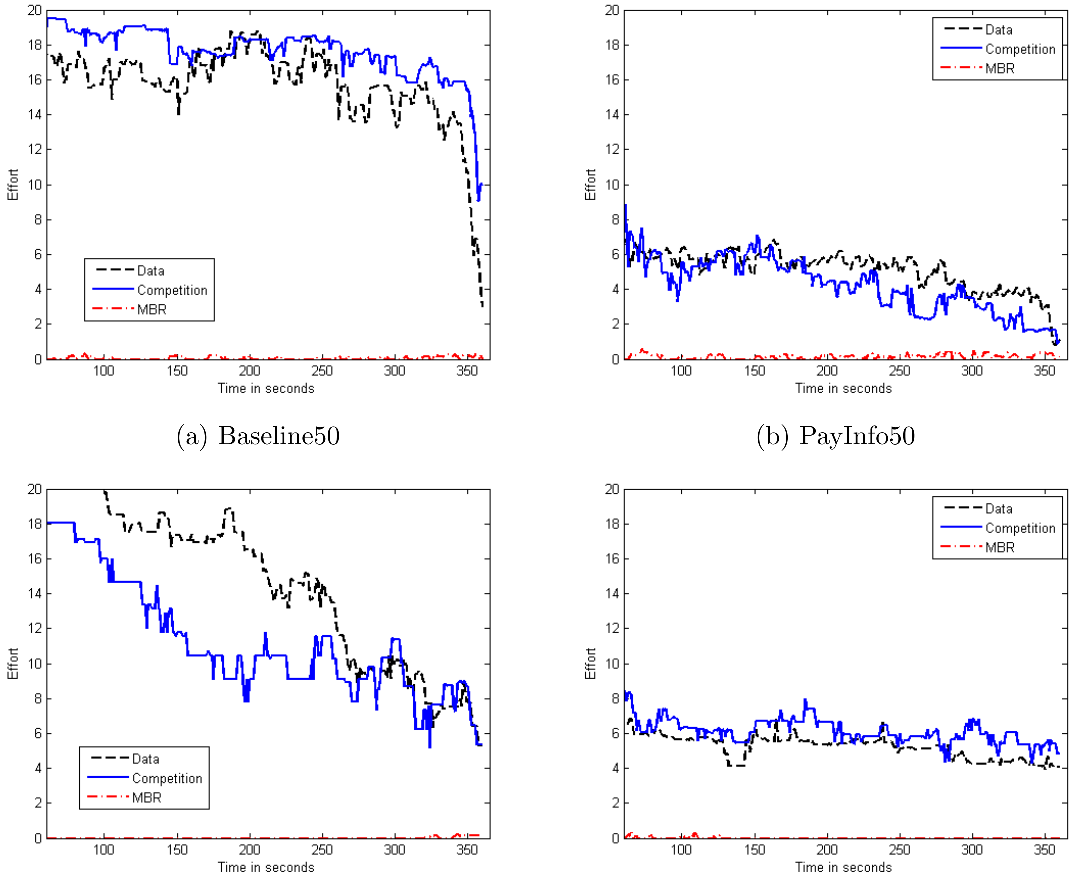


(c) Baseline8



(d) PayInfo8

Figure 21: Fitting effort dynamics with learning rules: most connected individual



(a) Baseline50

(b) PayInfo50

(c) Baseline100

(d) PayInfo100

Figure 22: Fitting effort dynamics with learning rules: most connected individual (cont.)

7 Conclusion

There is a large body of research that describes the structure of large empirical networks. A recurring theme in this work is that networks exhibit great inequality in connections. The economic theory of network formation shows that the trade-off between the costs of linking and the benefits of direct and indirect links is resolved in strategic models in favor of unequal networks. However, experiments on these models show that subjects do not form such networks. This mismatch between the theory and the experimental evidence provides the motivation for our paper.

We develop a new platform for the study of network formation. The platform allows for continuous time linking and effort choice and it allows for large scale experiments (up to 100 subjects). The paper presents an experiment on this platform; we test the predictions on specialization on linking and efforts in the model of Galeotti and Goyal (2010).

Our experiments provide strong support for the specialization prediction. Moreover, and in line with the theory, the specialization is more pronounced in larger groups. The second finding is that scale interacts powerfully with provision of information on others's payoffs. In the treatment where subjects see only their own payoffs as group size grows, the most connected individuals compete fiercely—they exert large efforts and have small earnings. By contrast, when a subject sees everyone's payoffs, as groups size grows, there is limited competition among highly connected subjects—they exert little effort and have large earnings. In the former setting, subjects always pick the pure influencer outcome, while in the latter case they often pick the pure connector outcome. We show that the treatment effects on effort dynamics can be reconciled with the individual learning rule that combines myopic best response and a desire to have many connections.

References

J. Andreoni and J. Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

O. Baetz. Social activity and network formation. *Theoretical Economics*, 10:315340, 2015.

V. Bala and S. Goyal. A non-cooperative model of network formation. *Econometrica*, 68 (5):1181–1231, 2000.

A.-L. Barabsi and R. Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999.

S. K. Berninghaus, K.-M. Ehrhart, and M. Ott. A network experiment in continuous time: The influence of link costs. *Experimental Economics*, 9(3):237251, 2006.

G. E. Bolton and A. Ockenfels. A theory of equity, reciprocity and competition. *America Economic Review*, 100:166–193, 2000.

M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

Y. Bramoulle and R. Kranton. Public goods in networks. *Journal of Economic Theory*, 135:478–494, 2007.

Y. Bramoulle, A. Galeotti, and B. Rogers. *Oxford Handbook on Economics of Networks*. Oxford University Press, 2016.

E. Calford and R. Oprea. Continuity, inertia, and strategic uncertainty: A test of the theory of continuous time games. *Econometrica*, 85:915–935, 2017.

C. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, 2003.

C. Camerer and T. H. Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67:827–874, 1999.

C. Camerer, T. H. Ho, and J.-K. Chong. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*, 104: 137–188, 2002.

A. G. Chandrasekhar, H. Larreguy, and J. P. Xandri. Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, 2019.

G. Charness, M. Corominas-Bosch, and G. R. Frechette. Bargaining and network structure: An experiment. *Journal of Economic Theory*, 136:28–65, 2007.

G. Charness, F. Feri, M. A. Meléndez-Jiménez, and M. Sutter. Experimental games on networks: Underpinnings of behavior and equilibrium selection. *Econometrica*, 82(5): 1615–1670, 2014.

G. B. Charness and M. Rabin. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117:817–869, 2002.

P. Eades. A heuristic for graph drawing. *Congressus numerantium*, 42:149–160, 1984.

A. Falk and M. Kosfeld. Its all about connections: Evidence on network formation. *Review of Network Economics*, 11:Article 2, 2012.

E. Fehr and K. M. Schmidt. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114:817–868, 1999.

D. Friedman and R. Oprea. A continuous dilemma. *American Economic Review*, 102(1): 337–63, 2012.

T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

A. Galeotti and S. Goyal. The law of the few. *The American Economic Review*, 100(4): 1468–1492, 2010.

J. K. Goeree, A. Riedl, and A. Ule. In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67(2):445–466, 2009.

S. Goyal, J. Moraga, and M. van der Leij. Economics: an emerging small world? *Journal of Political Economy*, 114(2):403–412, 2006.

S. Goyal, S. Rosenkranz, U. Weitzel, and V. Buskens. Information acquisition and exchange in social networks. *Economic Journal*, 127:2302–2331, 2017.

B. Herskovic and J. Ramos. Acquiring information through peers. *American Economic Review*, 110:2128–52, 2020.

D. Hojman and A. Szeidl. Core and periphery in networks. *Journal of Economic Theory*, 139:295–309, 2008.

C. A. Holt and S. K. Laury. Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655, 2002.

Y. Hu. Efficient, high-quality force-directed graph drawing. *Mathematica Journal*, 10(1): 37–71, 2005.

S. Huck, H.-T. Normann, and J. Oechssler. Learning in cournot oligopoly-an experiment. *Economic Journal*, 109:C80–C95, 1999.

R. M. Isaac and J. M. Walker. Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 103(1):179–199, 1988.

M. Jackson and A. Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74, 1996.

M. O. Jackson and B. W. Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, 2007.

M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6), 2014.

J. H. Kagel and D. Levin. The winner's curse and public information in common value auctions. *The American economic review*, pages 894–920, 1986.

M. Kearns, S. Judd, and Y. Vorobeychik. Behavioral experiments on a network formation game. *ACM EC 2012*, 2012.

J. O. Ledyard. Public goods: A survey of experimental research. *Working Paper, University Library of Munich,*, 1994.

S. Leider, M. Mobius, T. Rosenblat, and Q.-A. Do. Directed altruism and enforced reciprocity in social network. *Quarterly Journal of Economics*, 124(1):1815–1851, 2009.

J. Perego and S. Yuksel. Searching for information. *Working Paper Columbia University*, 2016.

J. Pettit, D. Friedman, C. Kephart, and R. Oprea. Software for continous game experiments. *Experimental Economics*, 17:631–648, 2014.

B. Rammstedt and O. P. John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212, 2007.

A. Riedl, I. M. Rohde, and M. Strobel. Efficient coordination in weakest-link games. *The Review of Economic Studies*, 83(2):737–767, 2016.

K. Schlag. Why imitate, nd if so, how? a bounded rational approach to multi-armed bandits. *Journal of Economic Theory*, 78:130–156, 1998.

B. van Leeuwen, T. Offerman, and A. Schram. Competition for Status Creates Superstars: an Experiment on Public Good Provision and Network Formation. *Journal of the European Economic Association*, 2019.

ONLINE APPENDICES

A Theory

The following proposition shows that, under discrete values of personal effort, a sufficiently high cost of linking implies a pure influencer equilibrium (for any group size n) and an approximate pure connector equilibrium (for a sufficiently large group size).

Proposition 2. *Suppose payoffs are given by (1), $a_1 = 1$, $a_2 \in (0,1)$, $\hat{y} \in X = \{0, 1, 2, \ldots, \overline{x}\}$, and $c(\hat{y} + a_2 - 1) < k < c\hat{y}$.*

(a) Every Nash equilibrium $s^* = (x^*, g^*)$ is such that g^* is a periphery sponsored star structure where the hub is a pure influencer investing $\hat{y}$ and every spoke invests 0.

(b) If $n \geq 2 + \frac{\hat{y}-1}{a_2}$, there exist pure connector ϵ-equilibria where the hub invests 0 while m spokes invest 1 and others invest 0 (with m s.t. $(m-1)a_2 < \hat{y} \leq 1 + (m-1)a_2$).

Proof. It follows from Proposition 1 that the pure influencer equilibrium always hold, regardless of n. Moreover, the pure connector equilibrium holds only if $n \geq 2 + \frac{k}{a_2(c\hat{y}-k)}$, in which case it requires every spoke to personally invest $\frac{\hat{y}}{1+(n-2)a_2}$. Since $c(\hat{y} + a_2 - 1) < k$ implies $\hat{y} < \frac{k}{c} + 1$, we have that $n \geq 2 + \frac{k}{a_2(c\hat{y}-k)} > 2 + \frac{\hat{y}-1}{a_2}$, and consequently $\frac{\hat{y}}{1+(n-2)a_2} < 1$ for any $n \geq 2 + \frac{\hat{y}-1}{a_2}$. However, since the lowest positive effort that can be made in the game is 1, no more than $m = \frac{\hat{y}-1}{a_2} + 2$ players (with $m < n$) can benefit from making such minimum positive effort. In this case, each of those positive investors accesses $(m-1)a_2$ from forming a link and therefore earns $f(1 + (m-1)a_2) - c - k$, which is strictly less than if they unilaterally deviate by forming no link and investing $\hat{y}$ since $c(\hat{y} + a_2 - 1) < k$ can be rewritten as $(m-1)a_2 < \frac{k}{c}$. As a result, the pure connector outcome is an ϵ-equilibrium whenever $\epsilon > f(\hat{y}) - f(1 + (m-1)a_2) - c(\hat{y} - 1)$ where m is the number of investing spokes such that $(m-1)a_2 < \hat{y} \leq 1 + (m-1)a_2$. $\square$

B Experimental platform

B.1 Network visualization

The force-directed algorithms of the network visualization tool use attraction and repulsion forces between nodes in the network and gravity force toward the center of the screen, in order to readjust their positions in two-dimensional space and improve the overall visibility on the subjects' screen.

Any two nodes o and o' in the network repulse each other with a *repulsion force* $F_r(o, o')$ in order to avoid overlaps and allow a sparse visualization of the network. It is modelled as a decreasing function of the Euclidean distance between two nodes $dist(o, o')$, implying that close nodes repulse more than distant nodes. Two connected nodes o and o' in the network apply an *attractive force* $F_a(o, o')$ towards each other to allow for visual proximity. A classical approach of modelling attraction force is a linear and positive relation with the distance, implying that close nodes attract less than distant nodes. Finally, every node o applies a *gravity force* $F_g(o)$ to the center of the spatialization space O to pull the entire network towards the center of the screen. In particular, such a force allows disconnected components to be within reasonable distance from each other, and therefore more easily visualized on the screen.

The net force vector applied to any node o resulting from the above three forces is then given by the following form of weighted sum (where F_x and F_y represent corresponding force vectors applied to the x and y axes of the Euclidean space respectively):

$$F_x(o) = \frac{x_O - x_o}{dist(o, O)} F_g(o) + \sum_{o' \in N \setminus \{o\}} \frac{x_{o'} - x_o}{dist(o, o')} F_a(o, o') + \sum_{o'' \in N \setminus \{o\}} \frac{x_{o''} - x_o}{dist(o, o'')} F_r(o, o'') \quad (3)$$

$$F_y(o) = \frac{y_O - y_o}{dist(o, O)} F_g(o) + \sum_{o' \in N \setminus \{o\}} \frac{y_{o'} - y_o}{dist(o, o')} F_a(o, o') + \sum_{o'' \in N \setminus \{o\}} \frac{y_{o''} - y_o}{dist(o, o'')} F_r(o, o'') \quad (4)$$

Note that the computation of the repulsion force for every node can be a complex task, especially in the context of large networks. In order to address this issue, the experimental software approximates this computation using the well-known algorithm introduced by **?**. More concretely, it finds groupings of nodes in proximity and determines a repulsion force $F_r(o, c)$ between node o and the group of nodes with a center of mass c, in replacement of the brute force method of computing repulsion forces between all pairs of nodes. More details of this approximation algorithm are provided at the following website: `http://networks.econ.cam.ac.uk/net_formation/experiments.html`.

We turn back to Figure 6 to derive some intuition of how the net force equations aggregate forces for every node and the network is visualized in the two-dimensional space. The adaptive visualization in Figure 6b is obtained by using the force-directed algorithm. The network has a petal-like structure with three independent sub-components connected through a common player, P5. The visualization algorithm makes P5 to be located at the center of the screen because the neighbors of P5 repluse each other and surround P5, while each pair of P5's neighbors belonging to the same sub-component are in close proximity and positioned side by side. The three forces then operate to make the rest of players located to draw non-overlapping petal-like structures.

Dynamic adjustment. The above equations (3) and (4) describe the net forces that are applied for the visualization of the network, given the positions of all nodes and the links between nodes. When the network changes, the algorithm updates dynamically the network visualization by computing the corresponding velocity of nodes on both coordinate axes.

In order to get a sense of how the network visualization is updated, we turn again to the example of network visualization in Figure 6 and show how the algorithm makes the transition from the fixed visualization in Figure 6a to the adaptive visualization in Figure 6b. Six (slow-motion) snap shots of the transition are presented at the following website: `http://networks.econ.cam.ac.uk/net_formation/experiments.html`. They show how the hub player, P5, moves from the bottom of the fixed circle to the center of the screen, and the petal-like structures emerge. This dynamic adjustment occurs rapidly to arrive at Figure 6b.

In our large-scale experiment, this visualization tool improves graphical clarity of evolving networks and helps subjects distinguish between those who are more connected and those who are less connected. It is wothwhile to note that this tool allows interaction between the subject and the network: while the nodes are subject to the above attraction and repulsion forces, they can also be freely manipulated by the participant through the usual drag-select functionality. The creation and removal of links is also interactive through double-clicking on corresponding nodes. This network visualization tool is built on the open source Javascript library *vis.js.*

Model parameter setting used in the experiment:

- $K_g = -2000$
- $K_s = 0.04$
- $K_{cg} = 0.3$
- $L = 95$

- $D = 0.09$
- $T = 0.5$
- $V_{min} = 0.3$
- $V_{max} = 10$

B.2 Continuous time with asynchronous choices

Running the continuous time experiments in large groups poses a number of technical challenges. First, every action made by a subject on her computer must be updated instantly on the computer screens of all other participants through the server computer. Network visualization must also be correspondingly updated in real time. As the group size increases, the information flows across the computer network increases dramatically. This can cause communication congestion and lagged responses. Another challenge with a large scale experiment is that it is constrained by the limited capacity of existing laboratories. Large groups that cannot fit into a single lab therefore require remote interactions between subjects in different geographical locations (that is, across different labs). In order to handle both of these technical challenges, we use a Websocket protocol with enhanced two-way communication between the server and subjects' computers. It fits naturally into the environment of asychronous choices in real time and the updates are made only when necessary. Our Websocket technology relies on the Javascript run-time environment *Node.js.* [22]

C Experimental instructions

[In the following instructions, N is to be replaced with any value from $\{3, 7, 49, 99\}$ *depending on the treatment]*

Please read the following instructions carefully. **These instructions are the same for all the participants**. The instructions state everything you need to know in order to

[22] Since it only requires an internet connection and is compatible with most existing web browsers (e.g., Google Chrome, Mozilla Firefox, Internet Explorer), this technology makes no specific restriction on the physical location of every participant.

participate in the experiment. If you have any questions, please raise your hand. One of the experimenters will answer your question.

You can earn money by earning points during the experiment. The number of points that you earn depends on your own choices and the choices of other participants. At the end of the experiment, the total number of points that you have earned will be exchanged at the following exchange rate:

100 points = 1 Euro

The money you earn will be paid out in cash at the end of the experiment. The other participants will not see how much you earned.

Details of the experiment

The experiment consists of 6 (six) independent rounds of the same form. The first round is for practice and does not count for your payment. The next 5 rounds will be counted for your payment.

At the beginning of each round, you will be grouped with N other participants. This group will remain fixed throughout the 6 rounds. Each of the participants will be randomly assigned an identification number of the form "Px" where x is a number between 1 and N. Those numbers will be randomly changed across every round of the experiment. The actual identity of the participants will not be revealed to you during or after the experiment. The participants will always be represented as blue circles on the decision screen. You are always represented as a yellow circle identified as "ME".

Each round will last **6 (six) mins: the first minute will be a trial period, only the latter 5 minutes will be relevant for the earnings**. Your earnings in a given round will be based on everyones choice **at a randomly selected moment in the last 5 mins of the round**. In other words, any decision made before or after that randomly chosen moment will not be used to determine your points. This precise moment will be announced to everyone only at the end of the round, along with the corresponding behavior and earnings.

At the beginning of the experiment, you are given an initial balance of 500 points. Your final earnings at the end of the experiment will consist of the sum of points you earn across the 5 last rounds plus this initial capital (the first round will be used to familiarize yourself with the game and will have no influence on your earnings). Note that if your final earnings (i.e., the sum of your earnings across the 5 last rounds plus the initial endowment) go below 0, your final earnings will be simply treated as 0.

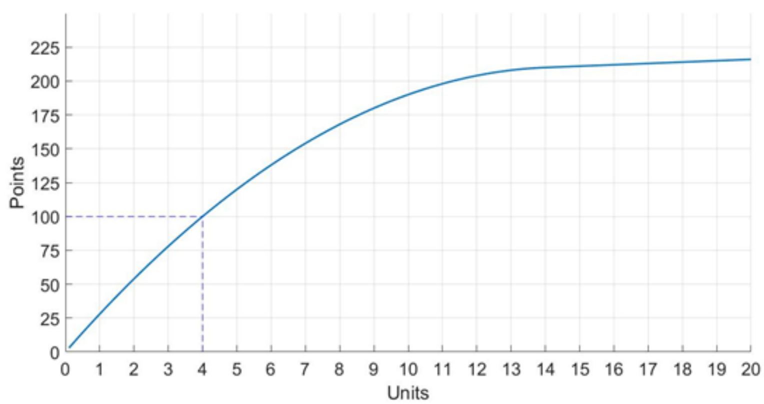
In each round, every participant will have choose two types of actions:

- **How many any units to buy/invest**: You may buy at most 20 units. Each unit costs you **11 points**.

- **Add/delete links with other participants**: You are linked with another person if you form a link with that person or that person forms a link with you (or both). You do not pay any fee for links formed by others. The people that you are linked with (regardless of whether you or they form the links) are called your neighbours. You automatically have access to **all units bought by your neighbours** as well as **half of the units bought by your neighbours' neighbours** (see below for an example). Each link you form costs you **95 points**.

You may revise your choices at any moment before the round ends. During a round, you will also be informed about every other participants most recent decision (units bought and formed links), which will be updated every 5 seconds or whenever you change your own choice.

At any moment, the total number of units you have access to (i.e., units you bought + units bought by your neighbours + units bought by your neighbours neighbours) generates points for you according to the following figure (for example, accessing 4 units generates 100 points, as shown by the dotted lines):



Moreover, having access to 20+m units generates 216+m points.

The computer screen will be split into two parts:

- **The middle side of the screen presents you and your local neighbourhood**. More precisely, you will see your neighbours, the neighbours of your neighbours, and the neighbours of neighbours neighbours. In other words, you will see the participants that are up to 3 links away from you.

- **The right side of the screen presents participants outside of your local neighbourhood**.

- **The left side of the screen presents the code for the players' net earnings in the network**. *[Payoff information treatment only]* The inner circle of each node from the middle or right part side of the screen is characterized by some color, which varies from **green** (high positive net payoff) to **red** (high negative net payoff) depending on the players corresponding net earnings.

Each node is described by their identification number "Px" and the number of units that they buy. Identification numbers "Px" are randomly assigned in every round. Therefore, every player is likely to have a different ID in different rounds. In the initial state of the network, nobody buys any unit and no link is formed.

Tutorial

Please follow this simple tutorial simulating a simple virtual scenario on the computer screen. In this tutorial you are interacting with 9 other players. In the initial state, you are not linked with anyone and you do not buy any units: you start at 0 points.

1. The slider allows you to choose how many units you wish to buy yourself. For example, buying 4 units costs you 44 points (= 4 units × 11 points, in red on the screen) and generates 100 points (according to the figure from the previous page, in green on the screen).

2. Initially, the nodes on the right side of the screen represent all other players (in this simulation, those players are not real people). The size of node reflects the total number of units bought by that node and the units accessed via the network. For example, P1-P4 are the largest nodes because these players have access to the most units.

3. You may choose to form a link with any player by simply double clicking on the corresponding node. For example, forming a link with P4 reveals that P1, P2, and P3 each form a link with P4, and P9 forms a link with P1. Forming a link with P4 costs you 95 points (in red on the screen), but it also gives you access to 8.5 units (7 from P4 + 0.5×1 from P1 + 0.5×1 from P2 + 0.5×1 from P3), which generates 174 points (according to the above figure, describing the benefit function in green on the screen). If you do not buy any additional unit yourself, your resulting net payoff is **79 points (= 174 points − 1 link × 95 points)**.

4. After forming a link with P4, you observe that some nodes remain unobserved (P5, P6, P7, and P8 on the right side). However, forming an additional link with P9 (by double clicking on the corresponding node) reveals that those nodes all form a link with P9. You were not allowed to observe them before because they were 4 nodes away from you (for example, P5 were connected to you via P4, P1, and P9). You can now observe them because they are only 2 nodes away from you (for example, P5 is connected to you via P9 only). Remember that you can only see players that are at most 3 nodes away. Assuming you still do not buy any unit yourself, your resulting net payoff is **16 points (= 206 points from accessing 12.5 units − 2 links × 95 points)**.

5. Alternatively, you may choose to remove a link that you previously formed by double clicking on the corresponding node. For example, after forming links with P4 and P9, removing the link with P4 leads to players P2 and P3 becoming unobserved again, as they are now more than 3 nodes away from you.

6. Note that varying the amount of units you buy directly affects the sizes of the nodes you are linked with as well as their neighbours. Indeed, the amount of units they each have access to includes the units you buy (the larger this amount, the larger the node).

7. You may also shape the visual structure of the network by dragging nodes as it pleases you.

Summary

Here is a brief description of information available on the decision screen:

1. The timer indicates elapsed time since the beginning of the round. Any round lasts **6 mins**. A moment will be randomly selected **in the last 5 mins** to determine everyones payoff. The time displayed will turn red when entering this interval.

2. **Only decisions made at the randomly selected moment in the round** matter to directly determine the earnings. The payoff may be negative at the end of a round. However, starting from a balance of 500 pts, any negative total of points at the end of the 5 rounds will be equivalent to 0 point.

3. The amount of units you have access is equal to the sum of **(1)** the units bought by you, **(2)** the units bought by your neighbours, and **(3)** half of the units bought by your neighbours neighbours.

4. You are represented as the yellow node, and your ID is "ME".

5. Every other nodes ID is represented as "Px" (inside the node) where x is a number. Every node has a unique ID, which is randomly reassigned in every round.

6. The size of each node determines **how many units that node has access to** (units bought personally plus units accessed from others, directly and indirectly).

7. The amount of units **bought personally by** a player is mentioned inside the corresponding node.

8. *[Payoff information treatment only]* The color of each node determines **that nodes net earnings** according to the code depicted on the left side of the screen.

D Network game interface

The decision making interface used in the experiment is similar across all treatments. More specifically, Figure 23 illustrates a (fictitious) example of a subject's computer screen in Treatment **Baseline100**. The top part of the screen depicts information about the timer indicating how much time has lapsed in the current round (the timer turns red when payoffs become effective, i.e., after more than 1 minute), the subject's own effort, which can be modified via the slider, and a comprehensive description of the subject's own payoff. Information about payoffs include gross earnings (output of function $f(.)$), the cost of effort (own effort multiplied by c), the cost of linking (number of links multiplied by k), and the net earnings (costs substracted from gross earnings). The bottom part of the screen shows detailed information about the network (the subject's node is highlighted in yellow): the subject's local network is represented on the left, other players outside of the subject's local network are found on the right. Note that a scrolldown feature is available for the subject to explore every player outside of his/her local network. Baseline treatments with smaller group sizes use the very same interface (the scrolldown feature is not available then because all players are then directly visible on the screen).

Similarly, Figure 24 illustrates a (fictitious) example of a subject's computer screen in Treatment **PayInfo100**. The only difference with the decision screen from Figure 23 is about the wider range of colors used to represent the border of each node depicted in the network. Any given node's color is directly associated with that node's corresponding payoff, according to the scale presented on the left part of the screen. payoff-information treatments with smaller group sizes use the very same interface.

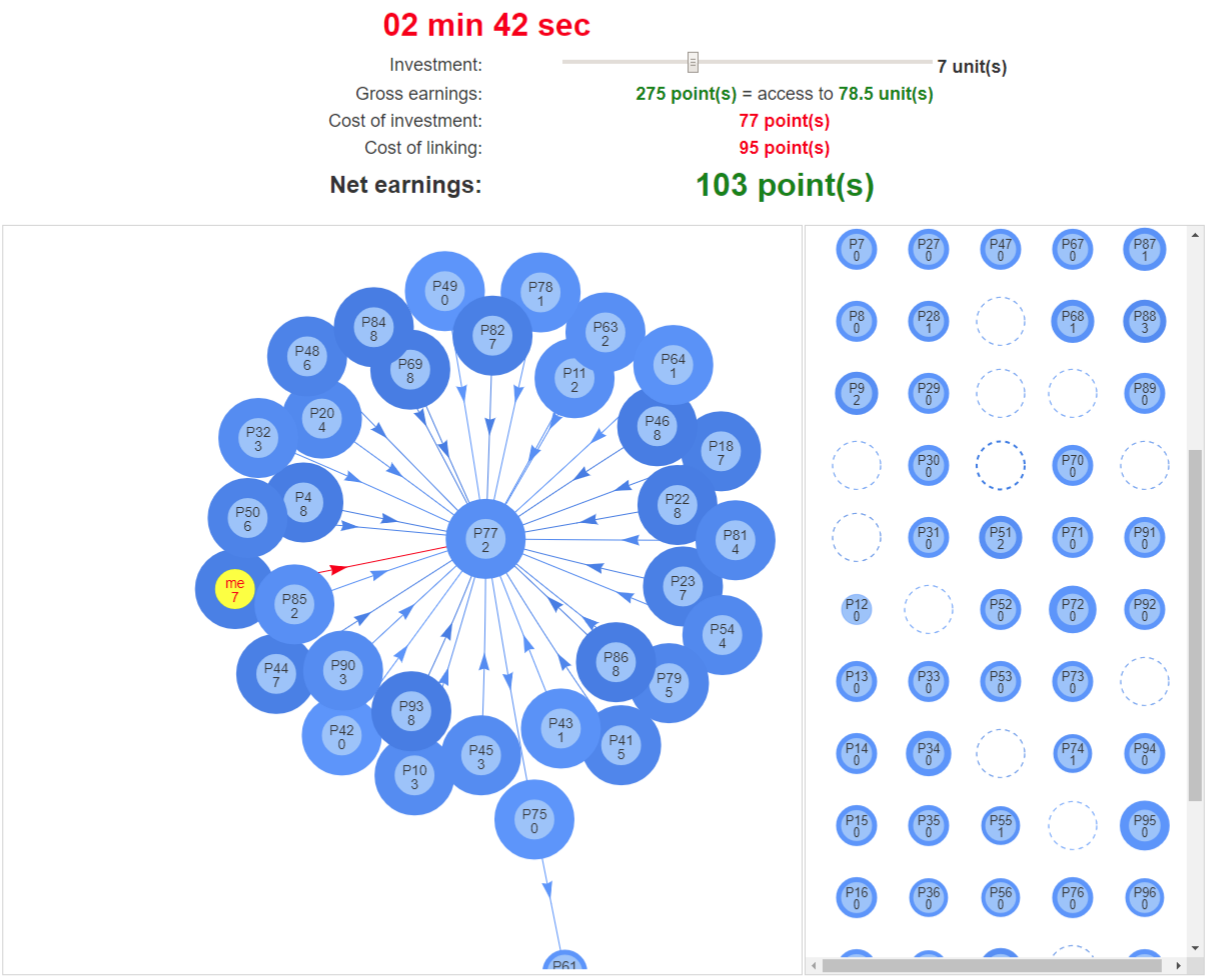


Figure 23: Example of decision screen for Treatment **Baseline100**

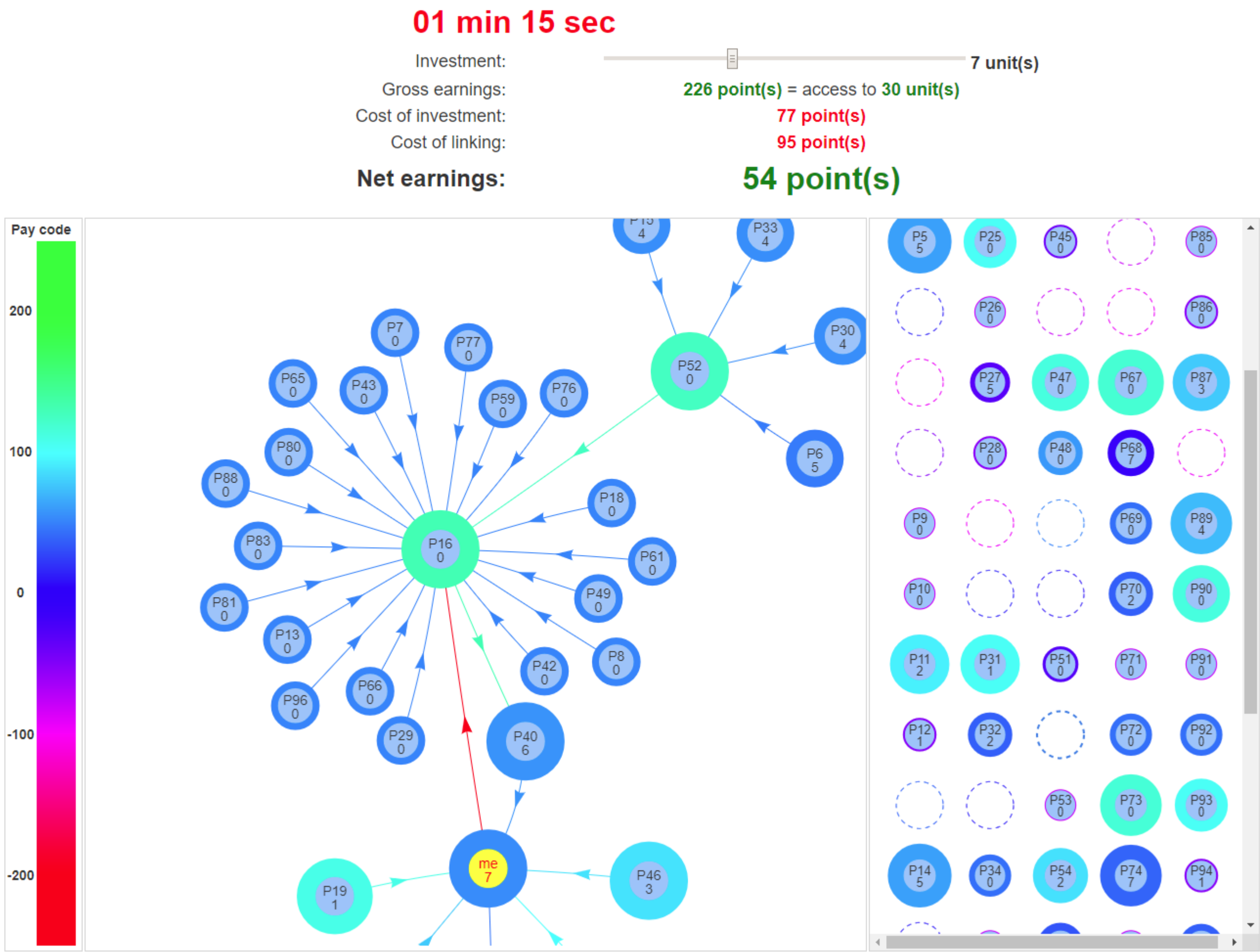


Figure 24: Example of decision screen for Treatment **PayInfo100**

E Questionnaires

At the end of the experiment, subjects answered a set of surveys aiming at measuring various types of individual differences. More precisely, incentivized measures of comprehension in network game, social preferences, and risk preferences were used. Finally non incentivized personality measures were used before which subjects filled up a debriefing questionnaire that includes demographics information.

E.1 Comprehension check

In order to assess the subjects' comprehension of the network game played during the experiment, we provided 5 questions, each of which with a unique correct answer. Each correct answer was rewarded with 0.1 euro for the subject.

The following first 2 questions were used across all treatments (correct answers are "11 pts" to question 1, and "95 pts" to question 2).

Question 1: In the previous game, how many points did investing one unit cost you?

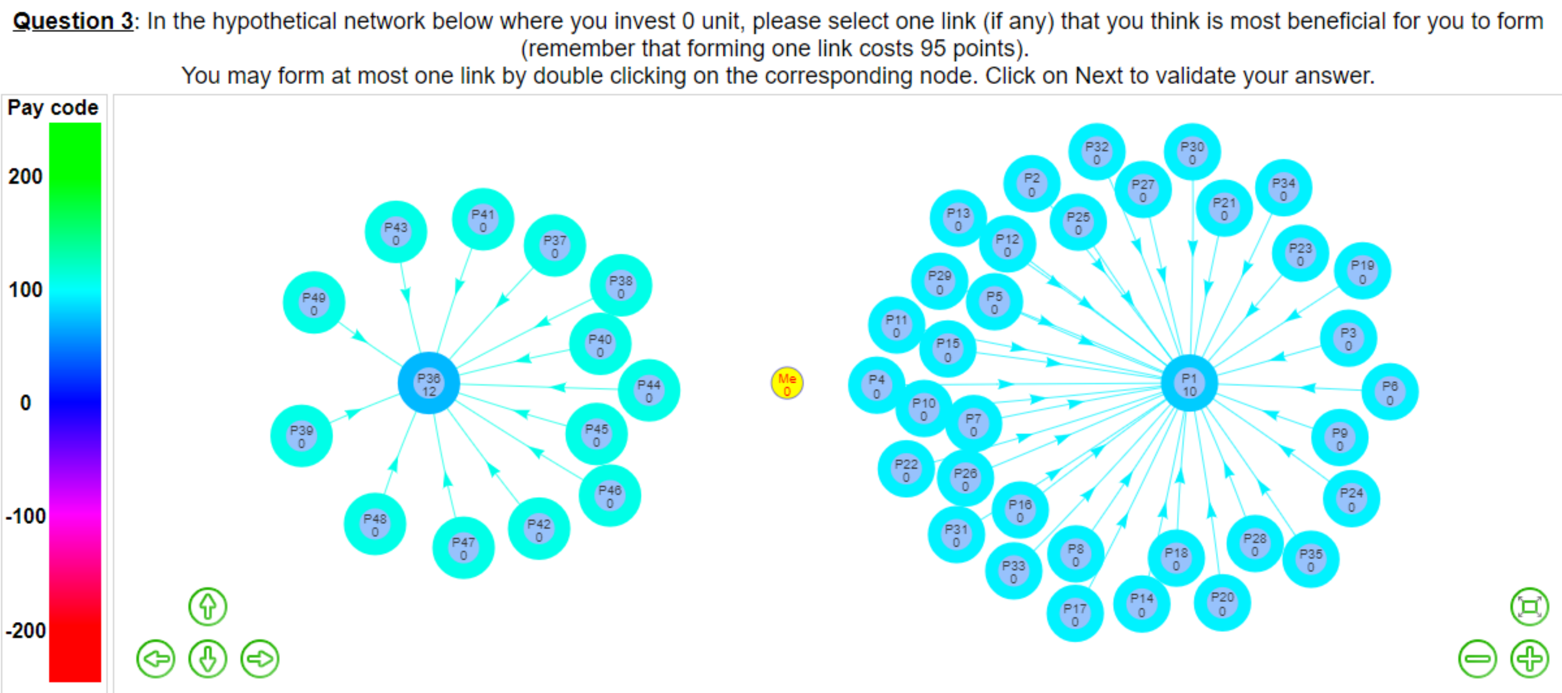
- 1 pts
- 11 pts
- 21 pts
- 31 pts
- 41 pts
- 51 pts
- more than 51 pts

Question 2: In the previous game, how many points did forming a link cost you?

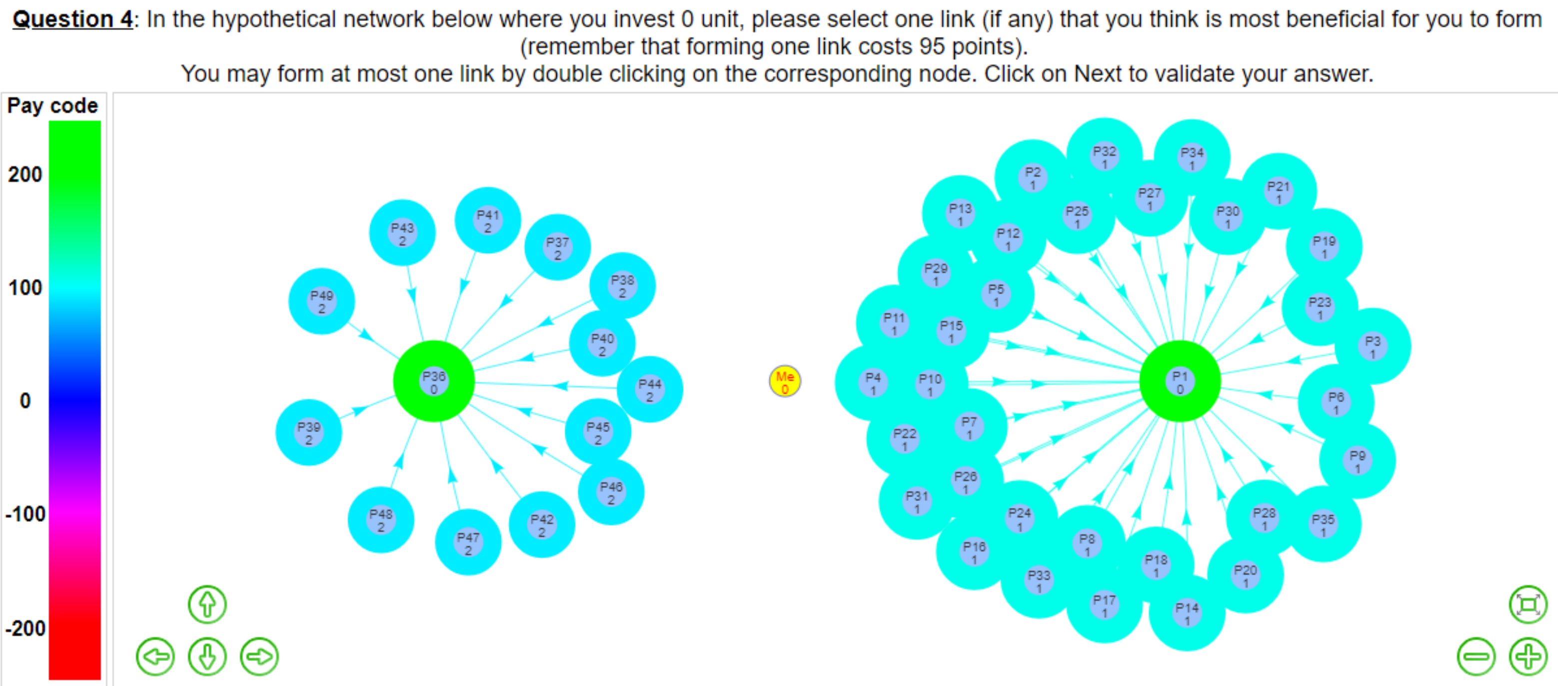
- 0 pt
- 25 pts
- 45 pts
- 65 pts
- 95 pts
- 115 pts
- more than 115 pts

The third question depicted below was used in the payoff information treatment with $n = 50$ (the correct answer is "P36"). This question was adapted in all other treatments by matching the number of nodes to the group size in the experiment, and by removing the colors in the baseline treatments.

The following questions 4 and 5 below were also used in the payoff information treatment with $n = 50$ (correct answers are "P1" for both questions 4 and 5). These questions were

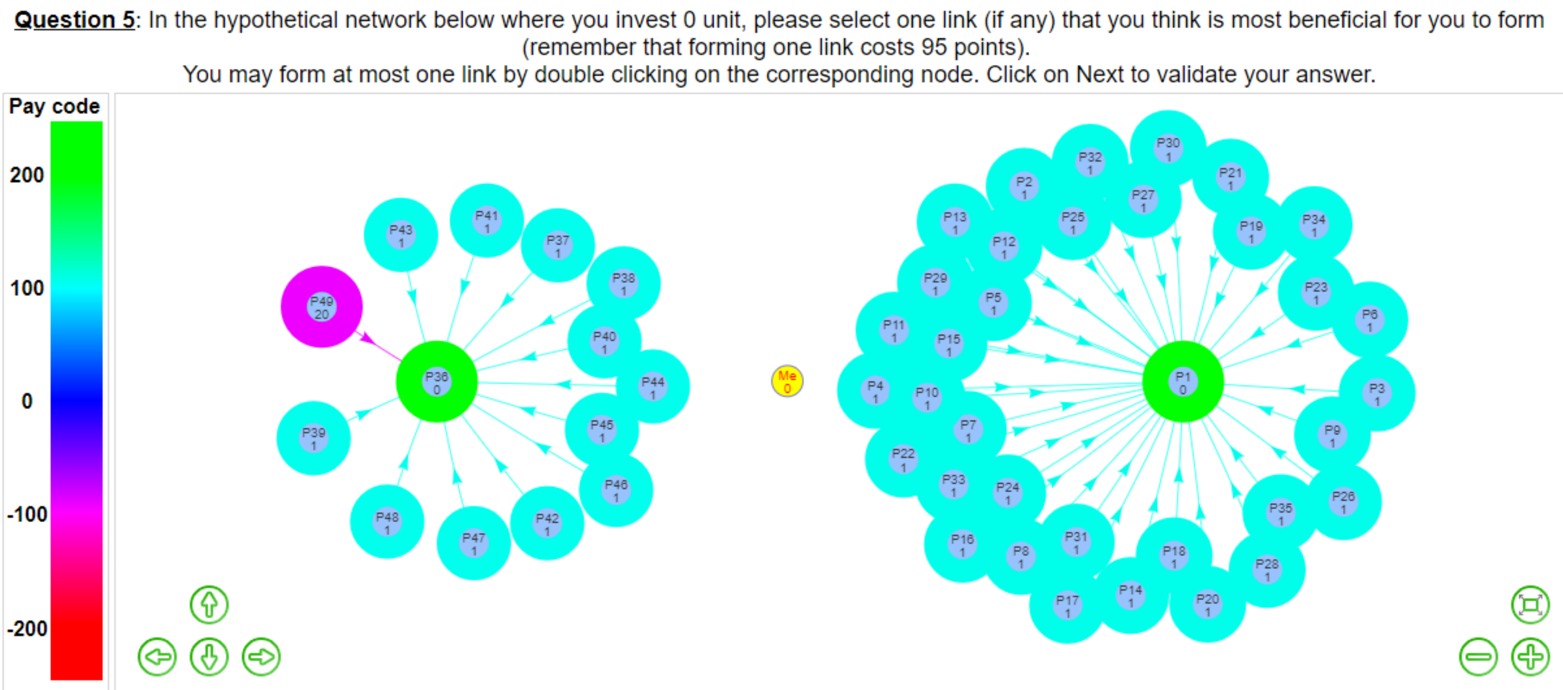


however adapted only in other treatments where $n > 4$ by again matching the number of nodes to the group size in the experiment. The reason for filtering the small group treatments (with $n = 4$) is that the limited number of nodes did not allow representing the corresponding scenarios. As before, these questions were also adapted to the baseline treatments by simply removing the colors to match the design of the actual game that subjects played.



E.2 Social preferences

The social preferences measure was adapted from Andreoni and Miller (2002) and involved a series of five money allocation tasks between the decision maker and some anonymous external participants of another experiment at the LINEEX lab (corresponding payments were therefore made to these external passive participants). The five tasks used in our experiment were represented through sliders as shown in the following figure:

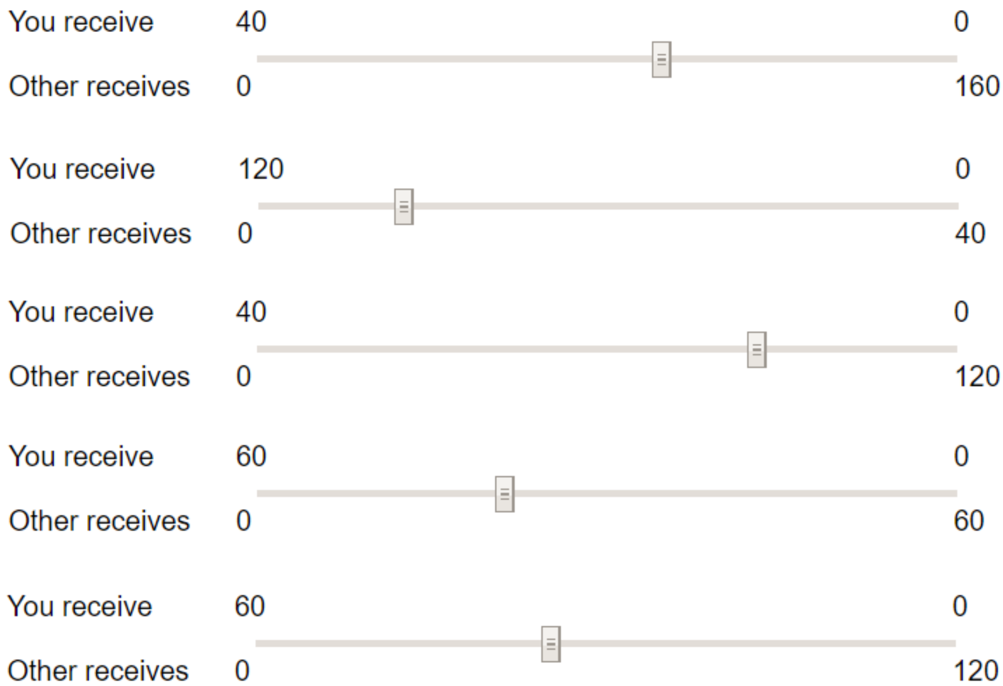


Note however that each question was presented in a different screen, and the order of presentation was randomized for every subject. Furthermore, 50 points were worth 1 euro both the subject, and the other anonymous external participant. Detailed instructions provided to the subjects, as well as a screenshot highlighting one of the above five questions are described below.

Instructions: You are asked to answer a series of 5 questions, each of which consists of selecting an allocation of points that you most prefer between yourself and an anonymous randomly selected person who is participating to a different experiment in this lab. At the end of the study, we will randomly select your allocation for 1 of the 5 questions to determine the payments for both you and the other person in this part. Your decisions will remain unknown to the other persons you are matched with.

E.3 Risk preferences

The risk preference measure was adapted from Holt and Laury (2002) and consisted of a series of five binary choices between lotteries, presented as in the figure below.



E.4 Personality test

Non incentivized measures were used through a simplified version of the Big Five personality inventory test adapted from Rammstedt and John (2007), as shown below.

Question 1

Please select your preferred allocation on the slider below
(values are in points, with 50 points = 1 euro):

You receive 17
Other receives 93

You receive	40	0
Other receives	0	160

Next

You are now asked to make 5 independent choices between two lotteries. According to **Lottery A**, you can win 2.00€ with a certain probability **p**, and 1.60€ otherwise. According to **Lottery B**, you can instead win 3.85€ with the same probability **p**, and 0.10€ otherwise. For each of the following 5 choices, which only differ in the value of the probability **p**, please select the lottery that you prefer. At the end of the study, we will randomly select one of your 5 preferred lotteries to determine your payment in this question.

	Lottery A			**Lottery B**
Choice 1:	2.00€ with probability 20/100, 1.60€ with probability 80/100	○	○	3.85€ with probability 20/100, 0.10€ with probability 80/100
Choice 2:	2.00€ with probability 35/100, 1.60€ with probability 65/100	○	○	3.85€ with probability 35/100, 0.10€ with probability 65/100
Choice 3:	2.00€ with probability 50/100, 1.60€ with probability 50/100	○	○	3.85€ with probability 50/100, 0.10€ with probability 50/100
Choice 4:	2.00€ with probability 65/100, 1.60€ with probability 35/100	○	○	3.85€ with probability 65/100, 0.10€ with probability 35/100
Choice 5:	2.00€ with probability 80/100, 1.60€ with probability 20/100	○	○	3.85€ with probability 80/100, 0.10€ with probability 20/100

Next

How well do the following statements describe your personality?

I see myself as someone who...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
1. ... is reserved					
2. ... is generally trusting					
3. ... tends to be lazy					
4. ... is relaxed, handles stress well					
5. ... has few artistic interests					
6. ... is outgoing, sociable					
7. ... tends to find fault with others					
8. ... does a thorough job					
9. ... gets nervous easily					
10. ... has an active imagination					

Next

F Appendix tables and figures

F.1 Regression tables

F.2 Appendix figures

Table 6: Regression analysis in the baseline treatments: time fraction

	Time fraction of being most connected (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	5.30***	5.29***		
	(0.50)	(0.50)		
Effort × Large group	0.84***	0.83***		
	(0.13)	(0.13)		
Indegree ratio (%) × Small group			-0.02	0.04
			(0.13)	(0.15)
Indegree ratio (%) × Large group			-1.10***	-1.15***
			(0.13)	(0.15)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.407	0.409	0.086	0.136

Notes: Robust standard errors, clustered by individual subject, are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 7: Scale effects on effort and payoff in the baseline treatments

	Mean effort			Median payoff		
	most connected	2nd most connected	others	most connected	2nd most connected	others
$N = 50$	6.61***	7.27***	0.32	-40.81***	-51.09**	28.82***
	(1.08)	(1.41)	(0.32)	(10.20)	(23.61)	(1.73)
Average in small group	8.77	5.24	2.65	86.50	81.00	85.00
Number of observations	60	60	1120	60	60	1120
R-squared	0.61	0.59	0.04	0.39	0.23	0.11

Notes: Robust standard errors (clustered by individual subject in the regression analysis of efforts) are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 8: Scale effects on effort and payoff in the baseline treatments

	Mean effort			Median payoff		
	most connected	2nd most connected	others	most connected	2nd most connected	others
$N = 100$	6.64***	11.06***	0.88***	16.54	-25.41*	53.20***
	(1.54)	(1.10)	(0.32)	(29.95)	(14.54)	(2.77)
Average in small group	8.77	5.24	2.65	86.50	81.00	85.00
Number of observations	55	55	1630	55	55	1630
R-squared	0.62	0.83	0.04	0.20	0.38	0.14

Notes: Robust standard errors (clustered by individual subject in the regression analysis of efforts) are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 9: Scale effects on outdegree in the baseline treatments

	Mean outdegree		
	most connected	2nd most connected	others
Large group	1.03***	0.75*	0.24***
	(0.35)	(0.40)	(0.05)
Average in small group	0.20	0.62	0.90
Number of observations	75	75	2590
R-squared	0.38	0.41	0.03

Notes: Robust standard errors, clustered by individual subject, are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 10: Regression analysis when information on others' payoff is observable: time fraction

	Time fraction of being most connected (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	6.20***	6.13***		
	(0.71)	(0.71)		
Effort × Large group	0.43***	0.43***		
	(0.15)	(0.15)		
Time fraction (%) × Small group			0.24**	0.28***
			(0.11)	(0.07)
Time fraction (%) × Large group			1.15***	1.13***
			(0.09)	(0.14)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.302	0.305	0.010	0.004

Notes: Robust standard errors, clustered by individual subject, are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 11: Treatment effects on effort and payoff

	Mean effort			Median payoff		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Payoff info	-1.35*	0.21	-0.04	0.76	-18.73***	-6.07**
	(0.76)	(0.65)	(0.37)	(8.94)	(4.04)	(2.38)
$N = 50$	6.53***	6.48***	0.30	-57.53***	-48.79**	29.99***
	(1.21)	(1.37)	(0.33)	(11.82)	(19.15)	(2.46)
Payoff info $\times$ $N = 50$	-8.90***	-5.72***	-0.39	142.27***	100.25***	-14.96***
	(1.83)	(1.93)	(0.44)	(18.28)	(29.02)	(2.26)
Mean or median in Baseline50	15.70	11.33	2.84	48.50	51.00	118.00
Number of observations	120	120	2240	120	120	2240
R-squared	0.54	0.38	0.04	0.21	0.14	0.11

Notes: Robust standard errors (clustered by individual subject in the regression of efforts) are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 12: Treatment effects on effort and payoff

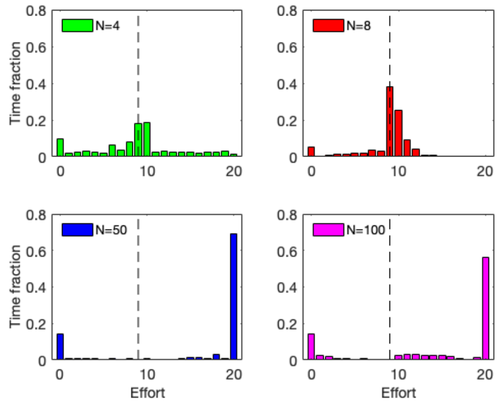
	Mean effort			Median payoff		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Payoff info	-0.55	0.20	0.06	-0.64	-15.63***	-11.36***
	(0.79)	(0.70)	(0.36)	(12.59)	(5.31)	(2.05)
$N = 100$	6.53***	10.26***	0.87***	40.61**	-43.22	51.44***
	(1.33)	(1.48)	(0.32)	(17.34)	(27.99)	(1.85)
Payoff info $\times$ $N = 100$	-9.82***	-12.48***	-1.28***	92.09	160.98***	-20.70***
	(1.65)	(2.01)	(0.41)	(150.87)	(29.02)	(2.67)
Mean or median in Baseline100	14.35	15.73	3.48	153.00	42.50	140.50
Number of observations	110	110	3260	110	110	3260
R-squared	0.53	0.68	0.07	0.07	0.26	0.13

Notes: Robust standard errors (clustered by individual subject in the regression of efforts) are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

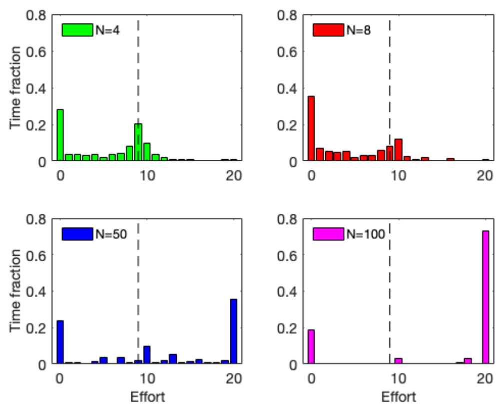
Table 13: Treatment effects on outdegree

	Mean outdegree		
	most connected	2nd most connected	others
Payoff info	-0.05	0.24	0.05
	(0.84)	(0.18)	(0.07)
Large group	1.14	0.94***	0.25***
	(0.93)	(0.33)	(0.05)
Payoff info × Large group	2.34	-0.80**	-0.01
	(2.42)	(0.39)	(0.07)
Average in large group baseline	1.90	1.32	1.12
Number of observations	150	150	5180
R-squared	0.43	0.34	0.04

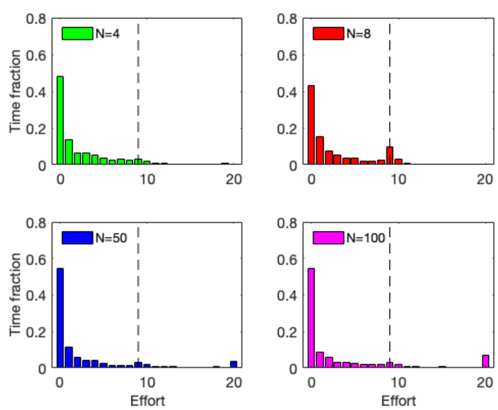
Notes: Robust standard errors, clustered by individual subject, are reported in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.



(a) the 1st most connected

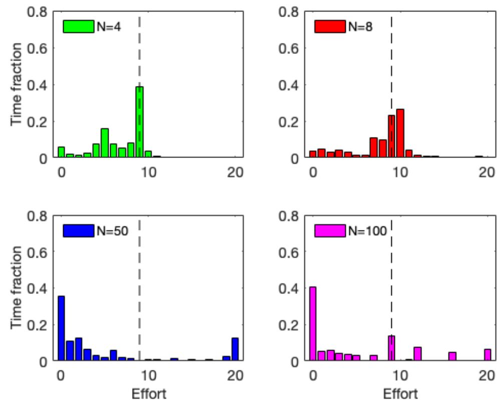


(b) the 2nd most connected

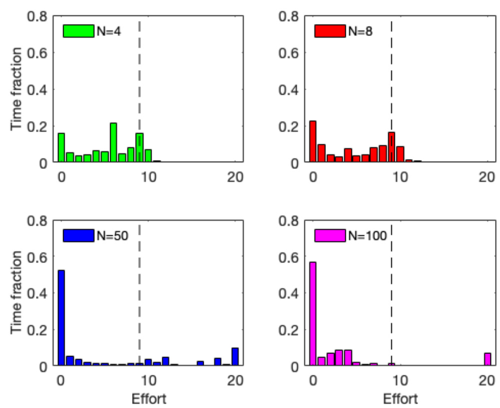


(c) the others

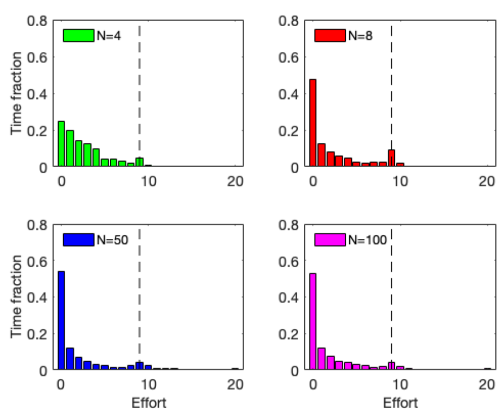
Figure 25: Distribution of efforts in the baseline treatment



(a) the 1st most connected

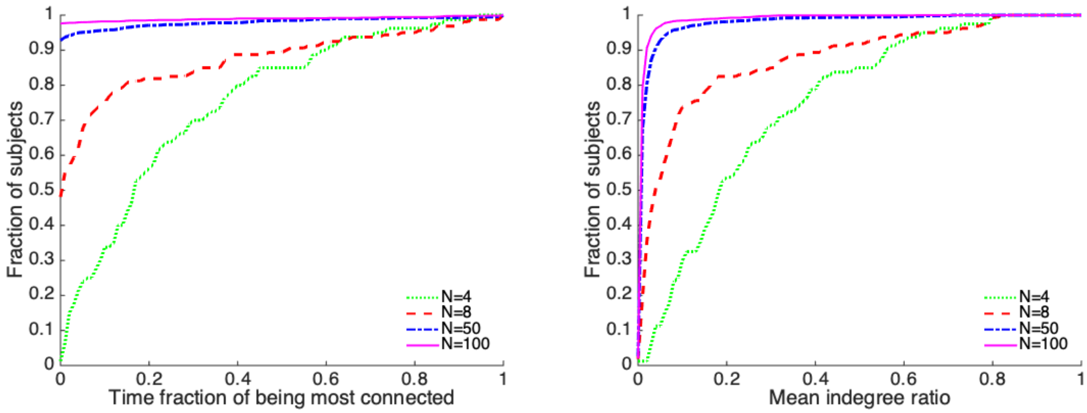


(b) the 2nd most connected



(c) the others

Figure 26: Distribution of efforts in the payoff information treatment



(a) Time fraction (b) Mean indegree ratio

Figure 27: Distribution of linking: information on others' payoff

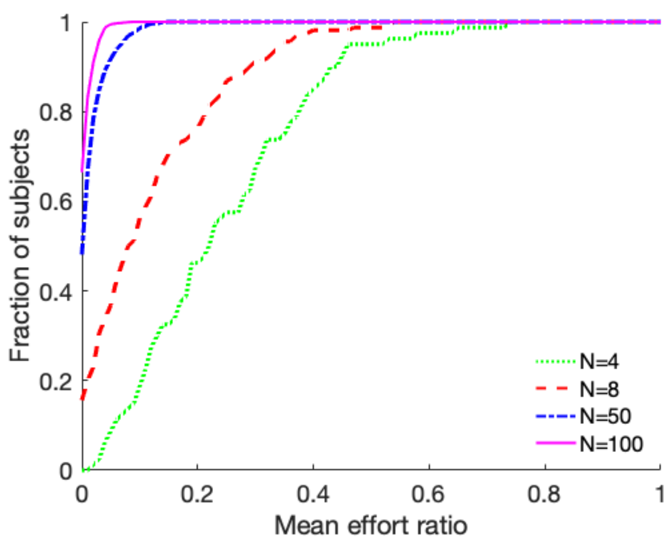
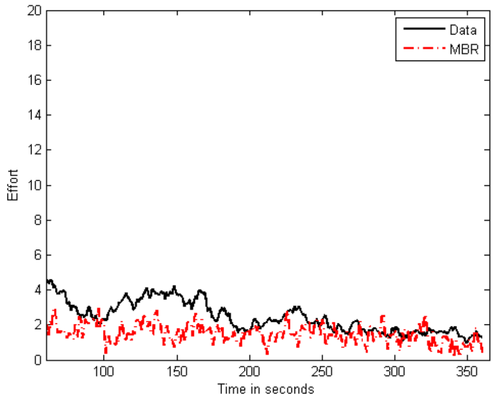
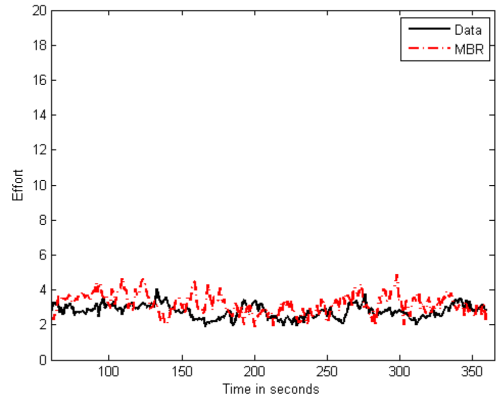


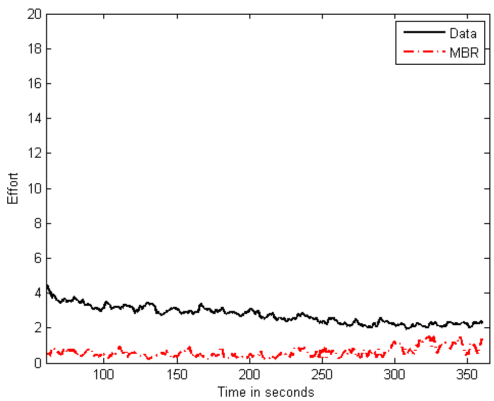
Figure 28: Distribution of Efforts in the payoff information treatments



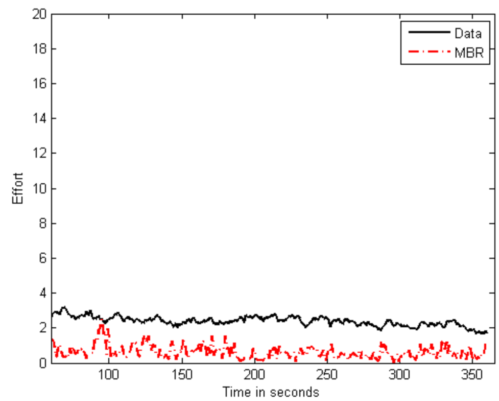
(a) Baseline4



(b) PayInfo4

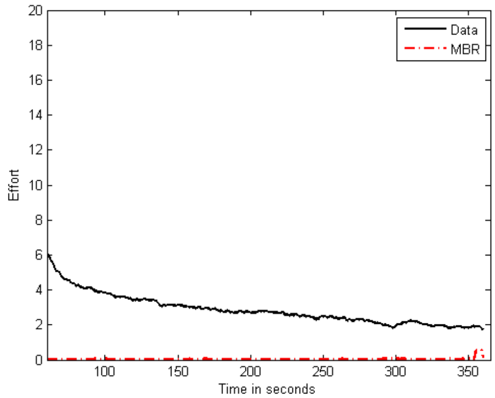


(c) Baseline8

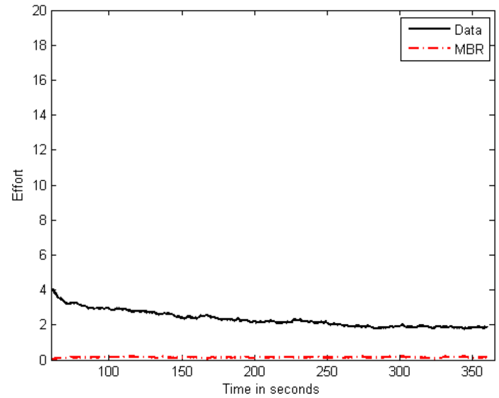


(d) PayInfo8

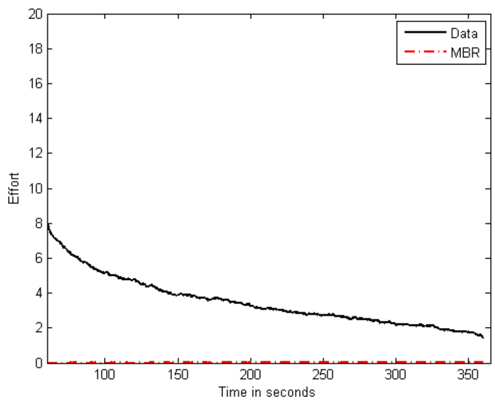
Figure 29: Fitting effort dynamics with learning rules: others



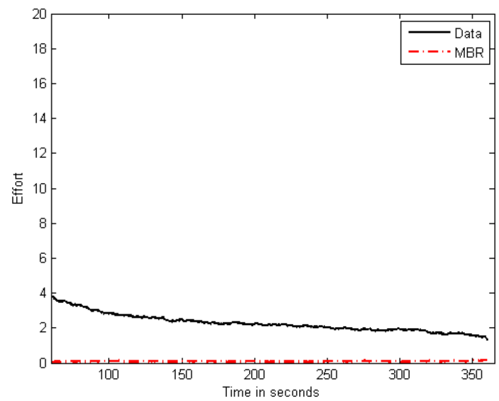
(a) Baseline50



(b) PayInfo50



(c) Baseline100

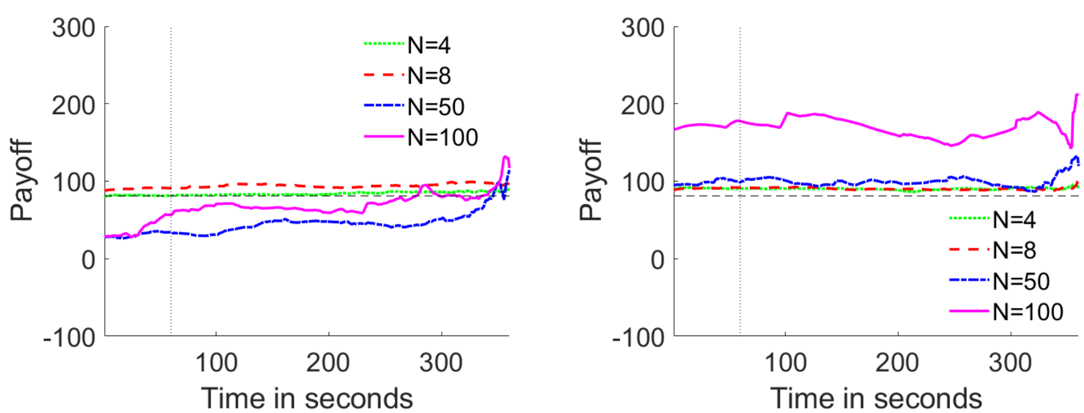


(d) PayInfo100

Figure 30: Fitting effort dynamics with learning rules: others (cont.)

	Baseline				Payoff information			
	$N = 4$	$N = 8$	$N = 50$	$N = 100$	$N = 4$	$N = 8$	$N = 50$	$N = 100$
10th percentile	48	16	9	36	33	33	7	1
25th percentile	48	51	68	96	33	61	59	61
Median	69	84	118	159	52	80	101	109
75th percentile	88	101	132	169	73	95	116	128
90th percentile	88	104	137	180	73	100	123	142
Equilibrium	85	85	85	85	85	85	85	85

Table 14: Distribution of Payoffs of Others: Last 2.5 minutes



(a) Baseline

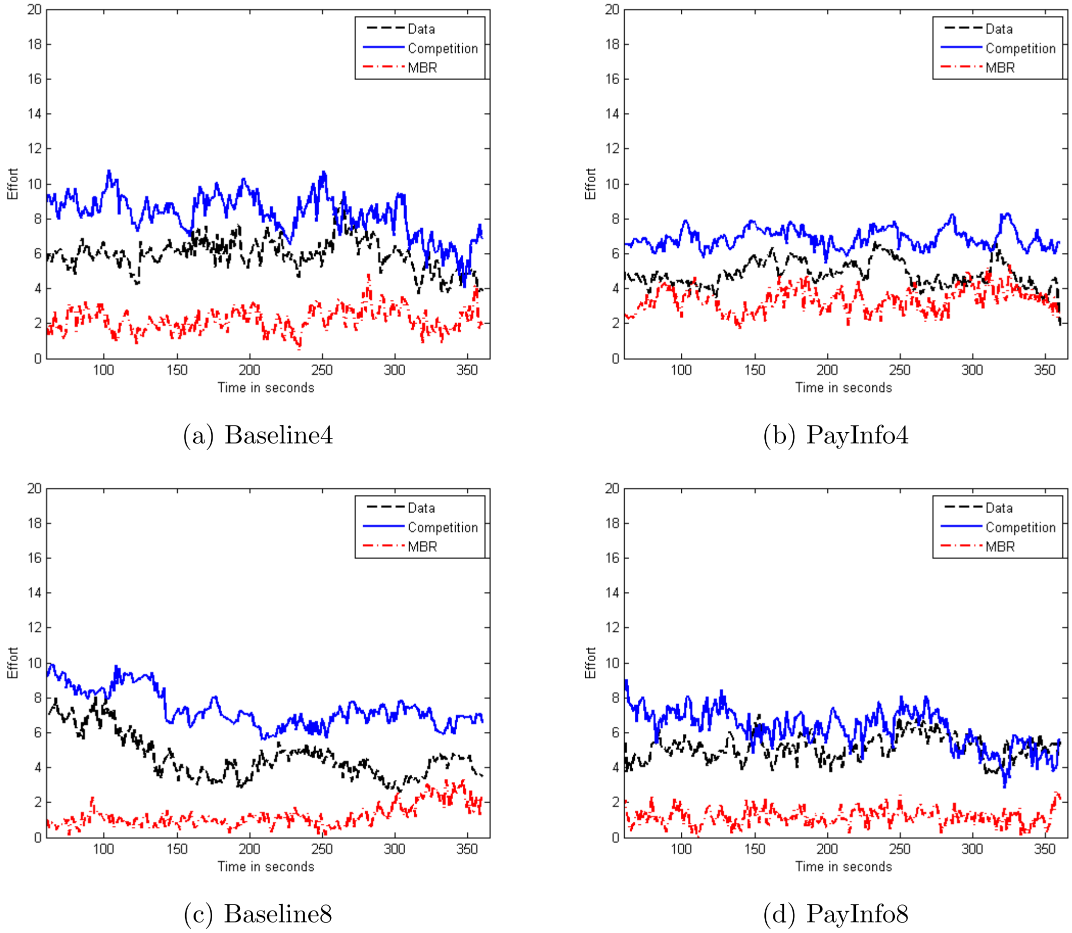
(b) PayInfo

Figure 31: Continuation payoffs

Table 15: Estimation of highly connected subjects' decision rules

	Base4	Pay4	Base8	Pay8	Base50	Pay50	Base100	Pay100
most connected subjects								
θ_i	3.0	2.9	6.4	6.4	44.0	8.4	50.6	40.2
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.32)	(0.00)	(0.65)
$\overline{x}_i$	10.3	7.5	9.6	8.8	18.7	13.3	16.8	9.3
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.21)	(0.14)	(0.17)
2nd most connected subjects								
θ_i	3.0	2.8	4.3	4.3	27.3	35.2	47.5	21
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.42)	(0.00)	(0.03)
$\overline{x}_i$	8.8	7.4	9.1	8.2	13.2	12.8	19.5	9.8
	(0.01)	(0.02)	(0.06)	(0.03)	(0.13)	(0.09)	(0.02)	(0.32)
SSR	376786	156671	306589	273231	724861	864740	1377068	459700

Notes: Bootstrapped standard errors (with 200 replications) are reported in parenthesis.



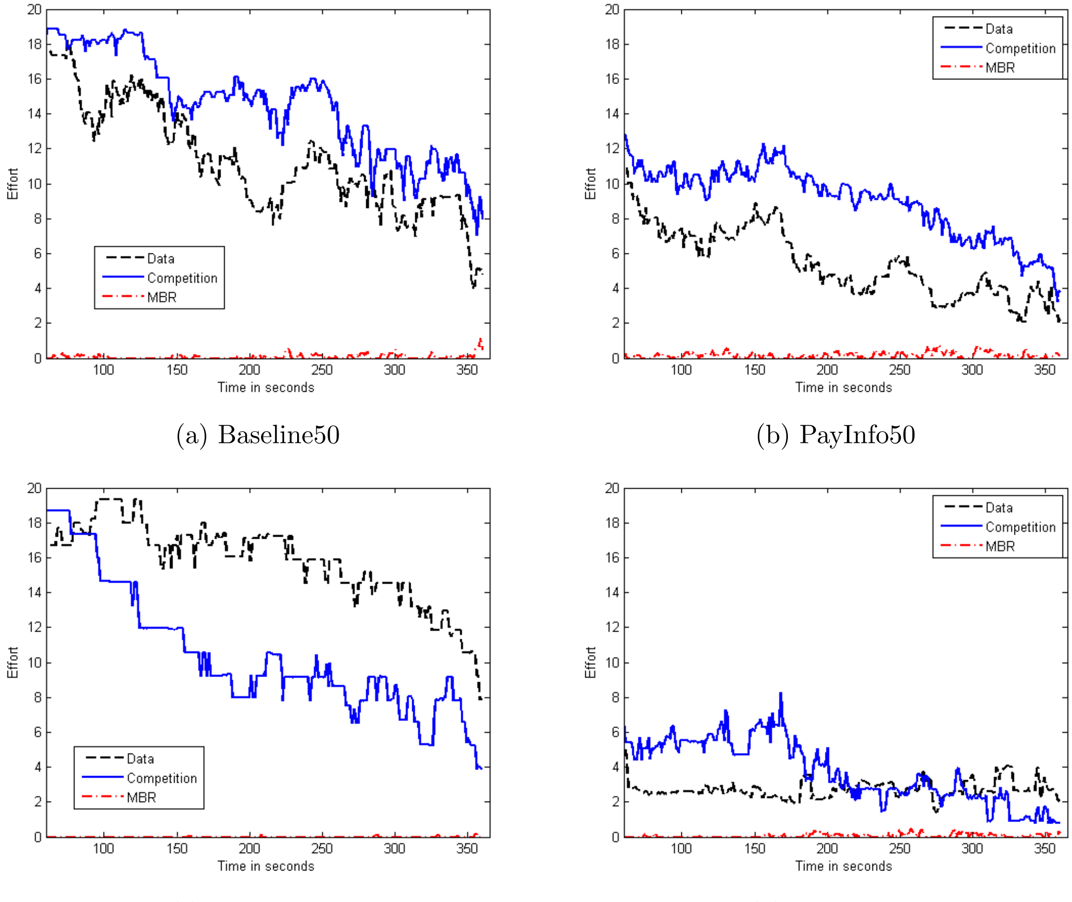
(a) Baseline4

(b) PayInfo4

(c) Baseline8

(d) PayInfo8

Figure 32: Fitting effort dynamics with learning rules: 2nd most connected



(a) Baseline50

(b) PayInfo50

(c) Baseline100

(d) PayInfo100

Figure 33: Fitting effort dynamics with learning rules: 2nd most connected (cont.)